

Sous la direction de
VÉRONIQUE GUÈVREMONT
COLETTE BRIN

INTELLIGENCE ARTIFICIELLE, CULTURE ET MÉDIAS



ÉTHIQUE IA
ET SOCIÉTÉ

Collection



ÉTHIQUE, IA ET SOCIÉTÉ

Collection dirigée par Lyse Langlois

Les innovations technologiques en intelligence artificielle sont un vecteur important du développement économique et peuvent contribuer significativement à l'amélioration des conditions de vie aujourd'hui et pour les générations futures. Leur impact soulève toutefois des enjeux majeurs, en matière d'emploi, de santé, d'éducation, de sécurité, de démocratie, de justice ou encore d'éthique. Ces enjeux traversent les frontières, nécessitent une réflexion et des actions pour lesquelles toute la communauté de recherche est appelée à jouer un rôle clé, et ce, en étroite collaboration avec les parties prenantes des milieux de l'industrie, du gouvernement et de la société civile. Cette collection présente des ouvrages érudits mais également des ouvrages de vulgarisation scientifique qui portent plus particulièrement sur les impacts sociétaux et éthiques de l'intelligence artificielle et du numérique.

Autres titres parus dans cette collection

Schallum Pierre et Fehmi Jaafar (dir.), *Médias sociaux : perspectives sur les défis liés à la cybersécurité, la gouvernamentalité algorithmique et l'intelligence artificielle*, 2020.

Karine Gentelet (dir.), *Les intelligences artificielles au prisme de la justice sociale / Considering Artificial Intelligence Through the Lens of Social Justice*, 2023.

Félix Pageau, Tenzin Wagmo et Emilian Mihailov (dir.), *Robots and Gadgets : Aging at Home*, 2024.

Cécile Petitgand, *Données personnelles : reprenons le pouvoir ! Réflexions sur la gouvernance citoyenne à l'ère numérique*, 2024, série Points de vue.

Intelligence artificielle, culture et médias

**Catalogage avant publication de Bibliothèque et Archives nationales
du Québec et Bibliothèque et Archives Canada**

Titre : Intelligence artificielle, culture et médias / sous la direction de Véronique Guèvremont et Colette Brin.

Noms : Guèvremont, Véronique, éditeur intellectuel. | Brin, Colette, éditeur intellectuel.

Description : Mention de collection : Éthique, IA et société | Comprend des références bibliographiques. | Comprend du texte en anglais.

Identifiants : Canadiana (livre imprimé) 20230084486 | Canadiana (livre numérique) 20230084494 | ISBN 9782763758770 | ISBN 9782763758787 (PDF)

Vedettes-matière : RVM : Intelligence artificielle—Aspect social. | RVM : Intelligence artificielle—Aspect moral. | RVM : Art génératif. | RVM : Algorithmes—Aspect social. | RVM : Médias et technologie.

Classification : LCC HM851.I57 2024 | CDD 303.48/33—dc23

Intelligence artificielle, culture et médias

Sous la direction de
Véronique Guèvremont
et **Colette Brin**



Presses de
l'Université Laval

Financé par le gouvernement du Canada
Funded by the Government of Canada



Nous remercions le Conseil des arts du Canada de son soutien.
We acknowledge the support of the Canada Council for the Arts.



Conseil des arts
du Canada

Canada Council
for the Arts

Les Presses de l'Université Laval reçoivent chaque année de la Société de développement des entreprises culturelles du Québec une aide financière pour l'ensemble de leur programme de publication.



Coordination de l'ouvrage: Iris Richer
Illustration de couverture: Refik Anadol
Révision linguistique: Solange Deschênes et Peter Frost
Mise en pages: Michèle Blondeau
Maquette de couverture: Laurie Patry

© Presses de l'Université Laval 2024

Tous droits réservés

Imprimé au Canada

Dépôt légal 2^e trimestre 2024

ISBN: 978-2-7637-5877-0

ISBN PDF: 9782763758787

Les Presses de l'Université Laval

www.pulaval.com

Intelligence artificielle, communication et médias sous la direction de
Véronique Guèvremont et Colette Brin © Les Presses de l'Université Laval
est mis à disposition selon les termes de la [licence Creative Commons
Attribution – Pas d'utilisation commerciale – Pas de modification 4.0
International.](#)



REMERCIEMENTS

Les directrices de l'ouvrage remercient l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA) ainsi que les membres du Comité scientifique du présent ouvrage pour leur expertise : René Audet (Université Laval), Sofian Audry (Université du Québec à Montréal, UQAM), Alain Kiyindou (Université Bordeaux Montaigne), Octavio Kulesz (éditeur et consultant pour l'UNESCO) et Sylvain Parasie (Sciences Po Paris). Elles remercient également Iris Richer (doctorante, Université Laval) pour la coordination de l'ouvrage, ainsi que les Presses de l'Université Laval (PUL).

TABLE DES MATIÈRES

INTRODUCTION

L'INTELLIGENCE ARTIFICIELLE, LA CULTURE ET LES MÉDIAS,
OU L'INTELLIGENCE ARTIFICIELLE À LA RENCONTRE
DU GÉNIE HUMAIN CRÉATEUR
VÉRONIQUE GUÈVREMONT ET COLETTE BRIN

Partie I – L'intelligence artificielle et la création artistique.....	4
Partie II – Le recours à l'intelligence artificielle et aux algorithmes de recommandation pour la diffusion des contenus journalistiques et culturels.....	5
Partie III – L'intelligence artificielle et la sauvegarde du patrimoine culturel	6
Partie IV – Les répercussions de l'intelligence artificielle sur la propriété intellectuelle, les droits humains et la diversité culturelle.....	7

PARTIE I

L'INTELLIGENCE ARTIFICIELLE ET LA CRÉATION ARTISTIQUE

1 - Surprise, interprétation, surgissement : penser la nouveauté à l'aune des machines créatives.....	13
FRANÇOIS LEVIN	
1. Le problème général : la machine créative	16
2. La position relationnelle	17
3. Une définition de la nouveauté insuffisante.....	23
4. La position processuelle	27

2 - Faire œuvre autonome ? Chomsky vs. Chomsky et autres œuvres produites par ou avec l'IA	35
CHRISTINE BLAIS ET OLIVIER BLAIS	
1. Étude de cas : <i>Chomsky vs Chomsky : First Encounter</i> (2020).....	37
2. Autres modèles d'IA créatives.....	44
3. La collaboration avec l'IA	47
4. Enjeux éthiques et économiques dans le marché de l'art : aller au-delà de la capacité créatrice de l'IA.....	49
3 - « Des artistes de silicium » ? IA et robots sur le devant de la scène : technicité et créativité dans le théâtre contemporain en Europe	55
MADELENA GONZALEZ ET CAMILLE HABAUT	
1. Robots, IA et cyborgs : contexte et dispositifs.....	57
2. Technicité et <i>poïétique</i> : l'IA au théâtre	62
3. « Artistes de silicium », robots performants et performatifs.....	66
Conclusion	69
4 - Émergence des dispositifs d'automatisation dans le jeu de société : le rôle de l'intelligence artificielle	73
SAMUEL FRANCBULU	
1. Contexte et enjeux de l'automatisation dans le jeu de société	75
2. L'IA analogique à l'épreuve des modèles sémio-communicationnels..	82
3. Perspectives pour la recherche-crédation	89
4. L'IA analogique : portée artistique, sociale et heuristique.....	94
Conclusion : l'IA ou l'affirmation du jeu comme produit culturel	98
5 - Intelligence artificielle, processus créatif et génération procédurale de jeux vidéo par conception automatisée.....	103
LAURA ISEUT LAFRANCE ST-MARTIN ET MAUDE BONENFANT	
1. Contexte entourant la conception automatisée de jeux vidéo	106
2. La conception automatisée de jeux vidéo à partir d'algorithmes d'intelligence artificielle	107
3. Définition du processus créatif dans le domaine de l'intelligence artificielle	110
4. Les enjeux de l'utilisation de l'intelligence artificielle dans la conception de jeux vidéo entiers	115
Conclusion	121

PARTIE II

LE RECOURS À L'INTELLIGENCE ARTIFICIELLE ET AUX ALGORITHMES DE RECOMMANDATION POUR LA DIFFUSION DES CONTENUS JOURNALISTIQUES ET CULTURELS

6 - Les outils liés à l'intelligence artificielle en journalisme en Amérique du Nord	129
NICOLAS ST-GERMAIN ET PATRICK WHITE	
1. Définitions de l'IA et son lien avec le journalisme	132
2. Intégration réussie des outils liés à l'IA	134
Conclusion	142
7 - Des usages des <i>Big Data</i>, des algorithmes et de l'intelligence artificielle par les plateformes audiovisuelles	149
ÉDOUARD GERMAIN, JUSTINE DORVAL ET ÉRIC GEORGE	
1. Les plateformes audiovisuelles internationales et leur appropriation des mégadonnées et de l'IA	152
2. De l'appropriation des mégadonnées et de l'IA par les services locaux	158
3. Éléments de réflexion pour aller au-delà... ..	161
8 - Recommandation algorithmique et dilemmes agentiels.....	167
LILY-CANNELLE MATHIEU	
1. Démarche méthodologique	170
2. Classification proposée : les composantes des dilemmes agentiels face à un système de recommandation culturelle algorithmique	173
3. Facteurs favorisant l'(in)action agentielle liée à la recommandation culturelle algorithmique	174
4. Facteurs freinant l'(in)action agentielle liée à la recommandation culturelle algorithmique	181
Conclusion	185

9 - Quand les abonnés et abonnées travaillent leur profil : Netflix, algorithmes et promesses de découvrabilité	193
ALIX BÉNISTANT, CHRISTINE CHEVRET-CASTELLANI ET SARAH LABELLE	
1. Promesses d'exploration dans une offre culturelle	197
2. Bricolage et travail du profil par les abonnés et abonnées	201
3. De l'appariement à la découvrabilité dans la consommation audiovisuelle	208
Conclusion	209
10 - Favoriser la visibilité du patrimoine cinématographique par les algorithmes : une expérimentation de recommandation de contenu à contenu appliquée au catalogue de la plateforme Tènk	213
ÈVE GIVOIS, ÉRIC KERGOSIEN, LAURE BOLKA-TABARY, SAMUEL GANTIER ET BERNARD JACQUEMIN	
1. Enjeux de la visibilité du cinéma patrimonial sur le Web	215
2. Un algorithme de recommandation humain ?	218
3. Expérimentation d'une recommandation automatisée	222
4. Discussion : limites de la démarche	235
Conclusion	238
11 - Les partenariats de données, une piste de solution collective aux enjeux de développement des applications de l'IA en art et en culture	243
VIÊT CAO ET JULIETTE DENIS	
1. Problématique	243
2. L'intelligence artificielle en arts et en culture	244
3. Les trois éléments d'une solution d'IA	244
4. Les partenariats de données, une piste de solution à la portée de tous	246
5. Présentation de quelques modèles de partenariats de données	249
6. Évolution et économie de la donnée et de l'IA	254
Conclusion	255

PARTIE III
L'INTELLIGENCE ARTIFICIELLE ET LA SAUVEGARDE
DU PATRIMOINE CULTUREL

12 - Intelligence artificielle et préservation des langues et pratiques culturelles sénégalaises : s'adapter ou disparaître dans la mondialisation ?	261
SERIGNE MOMAR SARR ET ALASSANE KORINGHO SAGNA	
1. Le Sénégal à l'ère de l'IA	264
2. Rapport entre culture, représentations sociales et IA chez certaines ethnies sénégalaises	268
3. Les défis des langues et des pratiques culturelles sénégalaises face à l'IA généralisée	271
Conclusion	276
 13 - Archives archéologiques, intelligence artificielle et sciences participatives : le projet « Bulliot, Bibracte et moi »	 279
EMMANUELLE PERRIN ET PHILIPPE CHASSIGNET	
1. Objectifs	280
2. Méthodologie : combiner intelligence artificielle et transcription participative	287
3. Résultats	292
 14 - Introduction de l'intelligence artificielle dans le métier de guide : quel impact sur la pratique de la visite guidée ?	 299
LINA UZLYTE	
1. Le métier de guide et l'intelligence artificielle	302
2. L'effet produit par l'introduction de l'intelligence artificielle sur la déontologie de la profession	309
Conclusion	321

PARTIE IV
LES RÉPERCUSSIONS DE L'INTELLIGENCE ARTIFICIELLE
SUR LA PROPRIÉTÉ INTELLECTUELLE, LES DROITS HUMAINS
ET LA DIVERSITÉ CULTURELLE

15 - IA et culture : de la création aux données	329
ALEXANDRA BENSAMOUN	
1. IA et création culturelle	331
2. IA et données de la culture.....	336
 16 - Between Device and Demiurge : AI and the Legal Status of AI Creations	 345
CAMILLE ALEGRE	
1. An Inherent Contradiction between Creativity and Automation.....	345
2. Anthropomorphic Bias and Arguments against the Legal Existence of AI Creators.....	348
3. A Creation without a Creator	352
4. The (Science) Fiction of Creative AI.....	357
 17 - Creative Piracy and AI Policing of Intellectual Property	 363
STEPHEN MONTEIRO	
1. Development of Online Piracy and Creative Production.....	365
2. AI Content Recognition and Piracy	367
3. Networked Creativity and Piracy	369
Conclusion	374
 18 - Algorithmes de recommandation : dispositif de la gouvernamentalité algorithmique menaçant les droits humains.....	 379
LUCIA FLORES ECHAIZ	
1. Algorithmes de recommandation.....	381
2. Gouvernamentalité algorithmique.....	384
3. Répercussions sur les droits humains.....	386
Conclusion	397

19 - Intégrer la diversité culturelle dans les instruments relatifs à l'encadrement de l'IA : vers une technologie culturellement responsable ?	405
IRIS RICHER ET CLÉMENCE VARIN	
1. La diversité culturelle et l'IA.....	407
2. La prise en compte de la diversité culturelle dans les instruments relatifs à l'IA.....	415
20 - Algorithmes et propagation de normes culturelles sexospécifiques.....	429
ELEONORE FOURNIER-TOMBS ET CÉLINE CASTETS-RENARD	
1. Propagation de normes discriminatoires envers les femmes dans des algorithmes biaisés	431
2. Des méthodes algorithmiques peuvent-elles réduire les hypothèses discriminatoires et les biais systémiques envers les femmes?.....	438
Conclusion	445
21 - L'IA dans la modération : censure des contenus LGBT sur TikTok et Twitter et terreau de nouvelles formes sémiotiques	449
THIBAUT GRISON ET VIRGINIE JULLIARD	
1. Du biais algorithmique à l'invisibilisation de communautés en ligne	451
2. La fabrique de stratégies contre la censure algorithmique	457
3. Pour une prise en compte des savoirs situés dans les études sur les discriminations algorithmiques	465
Notices biographiques des contributeurs et contributrices	471

INTRODUCTION

L'INTELLIGENCE ARTIFICIELLE, LA CULTURE ET LES MÉDIAS, OU L'INTELLIGENCE ARTIFICIELLE À LA RENCONTRE DU GÉNIE HUMAIN CRÉATEUR

Véronique Guèvremont¹ et Colette Brin²

Les systèmes d'intelligence artificielle (IA) ont des répercussions majeures sur l'ensemble de la chaîne de valeur d'innombrables secteurs. La culture et les médias n'y échappent pas. Dans ces domaines, comme partout ailleurs, « les technologies de l'IA peuvent rendre de grands services à l'humanité [...] mais [...] soulèvent également des préoccupations éthiques de fond³ ».

Biais, discrimination, inégalité, exclusion, censure et homogénéisation sont des risques fréquemment évoqués pour formuler des mises en garde d'ordre général, auxquels s'ajoutent des risques propres aux secteurs de la culture et des médias. Depuis plusieurs années déjà, les applications et les outils commerciaux – notamment les puissants algorithmes de recommandation auxquels ont recours les grandes plateformes de diffusion de contenus et les médias sociaux – inquiètent les défenseurs de la diversité culturelle et du pluralisme. Maintenant, ce sont aussi les outils

1. Professeure titulaire à la Faculté de droit de l'Université Laval.

2. Professeure titulaire au Département d'information et de communication de l'Université Laval.

3. *Recommandation sur l'éthique de l'intelligence artificielle*, Doc off UNESCO, Doc UNESCO SHS/BIO/PI/2021/1, 23 novembre 2021, 6^e alinéa.

conversationnels ou générateurs d'œuvres créées par IA qui soulèvent des interrogations : sur la nature de l'art, de la créativité, de l'écriture. Elles génèrent en outre des préoccupations nouvelles sur le rôle de l'artiste, du créateur, du journaliste. La perspective de son remplacement est évoquée⁴. La fascination engendrée par certaines œuvres primées, produites par l'IA ou avec son aide, ne contribue en rien à apaiser les esprits⁵.

Mais, à plusieurs égards, l'IA est également source d'occasions pour la culture et les médias. Écartant l'idée d'avènement de « robots artistes », certains préfèrent ainsi parler d'« artiste augmenté⁶ ». Loin de se limiter au créateur, l'IA stimule aussi le potentiel du producteur, distributeur et diffuseur. Ses bénéfices atteignent également le public. Quelques exemples suffisent à le démontrer.

L'IA transforme l'industrie du jeu vidéo, en créant des expériences immersives inédites, en plus des multiples emplois et des revenus faramenteux qu'elle génère⁷. Dans le domaine de la musique, l'IA accompagne les artistes dans l'exploration de nouveaux genres musicaux⁸. L'IA facilite par ailleurs la détection de violation des droits de propriété intellectuelle et certaines plateformes l'utilisent déjà à cette fin⁹. Il est de notoriété que

-
4. Julian Rothman et Shaina Feinberg, « Human Artists Take on their New Robot Competition », *New York Times*, 22 décembre 2023, <<https://www.nytimes.com/2022/12/23/business/ai-generated-illustration.html?searchResultPosition=12>>. Ojoma Ochain souligne que, « [d]’après certaines prévisions, l’IA produira une chanson du Top 40 d’ici 2027, générera une vidéo créative d’ici 2028 et écrira un best-seller de la liste du *New York Times* d’ici 2049 ». Voir : Ojoma Ochai, « Nouvelles opportunités et nouveaux défis pour des industries culturelles et créatives inclusives dans l’environnement numérique », dans *Re|penser les politiques en faveur de la créativité : la culture un bien public mondial*, Paris, UNESCO, 2022, p. 96.
 5. « L’intelligence artificielle, plus forte que l’Homme ? », telle est la question posée par un article publié en 2022 qui relate le débat houleux ayant suivi l’attribution d’un prix, lors de la Colorado State Fair, à une œuvre entièrement réalisée par Midjourney. Voir : « Une IA secoue le monde de l’art... en remportant un concours », *Arts in the City*, 7 septembre 2022, <<https://www.arts-in-the-city.com/2022/09/07/une-ia-secoue-le-monde-de-lart-en-remportant-un-concours/>>.
 6. Octavio Kulesz et Thierry Dutoit, « L’intelligence artificielle dans l’art et les industries culturelles et créatives. Panorama des technologies, expertises et bonnes pratiques dans l’espace francophone », Paris, OIF, 2020, p. 33, <https://www.francophonie.org/sites/default/files/2020-05/OIF_Guide-IA-art_VF.pdf>.
 7. « The Potential of Artificial Intelligence (AI) in Gaming », *ThinkML*, 7 avril 2023, <<https://thinkml.ai/the-potential-of-artificial-intelligence-ai-in-gaming/>>.
 8. Andrew R. Chow, « “There’s a Wide-Open Horizon of Possibility.” Musicians Are Using AI to Create Otherwise Impossible New Songs », *Time*, 5 février 2020, <<https://time.com/5774723/ai-music/>>.
 9. Par exemple, la plateforme d’écoute musicale Deezer. Voir sur ce point : Ricky Luliocci, « How Deezer’s AI is Safeguarding Musicians’ Intellectual Property Rights », *We Rave You Tech*, 25 juin 2023, <<https://weraveyou.com/tech/deezer-ai-detecting-copyright-infringement>>.

l'industrie du cinéma utilise abondamment l'IA pour produire ses effets spéciaux, mais elle permet désormais d'aller beaucoup plus loin, en créant des hypertrucages (*deepfakes*) ou en rajeunissant de plusieurs décennies des acteurs (on parle alors de *de-ageing*)¹⁰. Le spectacle vivant, les arts du cirque notamment, se tourne de plus en plus vers l'IA pour proposer des performances multidisciplinaires qui intègrent réalité virtuelle et augmentée. Grâce à l'IA, certaines performances intégrées au Métavers vont même jusqu'à générer une « émotion similaire » à celle qui est provoquée par la prestation en direct¹¹. Le domaine du patrimoine culturel n'est pas non plus en reste, l'IA transformant les façons de le transmettre et de le rendre plus facilement accessible¹². Finalement, le secteur des médias tire aussi profit des systèmes d'IA, par exemple pour la rédaction de comptes rendus ou de résumés dans les domaines financier et sportif, pour l'optimisation de la mise à disposition de textes selon le profil du lecteur, pour produire des titres plus accrocheurs, ou encore pour détecter des tendances, des nouvelles de dernière heure et même les fausses nouvelles¹³.

C'est donc sous de multiples angles que l'influence croissante de l'IA sur la culture et les médias mérite d'être examinée. D'où l'importance de stimuler la réflexion sur ces thèmes, généralement encore peu explorés par la doctrine. Les travaux de recherche sur l'IA dans les secteurs de la culture et des médias à partir des disciplines « mères », ainsi que des sciences humaines et sociales et du droit, sont d'ailleurs encore assez peu nombreux

-
10. Le rajeunissement de l'acteur Harrison Ford dans le film *Indiana Jones and the Dial of Destiny* a fait couler beaucoup d'encre. Voir notamment à ce sujet : David Smith, « "We're Going through a Big Revolution" : How AI is De-ageing Stars on Screen », *The Guardian*, 6 février 2023, <<https://www.theguardian.com/film/2023/feb/06/de-ageing-on-screen-ai-artificial-intelligence-film>>.
 11. Voir les propos d'Alexandre Teodoresco, directeur stratégie et développement à la Compagnie des 7 doigts, dans Diane Martin-Graser, « La révolution du Métavers : à quoi doit-on s'attendre ? », *CScience*, 3 décembre 2021 <<https://www.cscience.ca/2021/12/03/la-revolution-du-metavers-a-quoi-doit-on-sattendre/>>.
 12. Conseil de l'Europe, « Créer notre avenir : la créativité et le patrimoine culturel en tant que ressources stratégiques pour une Europe diversifiée et démocratique », Document d'information préparé par Arianna Traviglia, Conférence du Conseil de l'Europe des ministres de la Culture, Strasbourg, 1^{er} avril 2022, <<https://rm.coe.int/les-technologies-numeriques-y-compris-l-intelligence-artificielle-au-s/1680a5ea9c>>.
 13. Kati Bremme, « IA dans les médias. Cartographie des applications de l'Intelligence Artificielle dans les médias », *Méta-média*, 2019, p. 18-29, <<https://www.meta-media.fr/wp-content/uploads/sites/33/2019/06/Cartographie-IA-dans-les-medias-FR.pdf>> ; Sonali Verma, « Globe and Mail's Dynamic Paywall Increases Digital Subscriptions 51 % », *International News Media Association Ideas Blog*, 6 juillet 2021, <<https://www.inma.org/blogs/ideas/post.cfm/globe-and-mail-s-dynamic-paywall-increases-digital-subscriptions-51>> ; Gabe Bullard, « Smart Ways Journalists Can Exploit Artificial Intelligence », *Nieman Reports*, 20 juin 2023, <<https://niemanreports.org/articles/artificial-intelligence-newsrooms/>>.

et rarement mis en dialogue. Stimuler ces travaux est l'une des missions de l'Axe Arts, médias et diversité culturelle de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA), mission qu'il assume par diverses activités scientifiques¹⁴. La réalisation de cet ouvrage collectif en est l'une de ses manifestations.

L'appel à contributions pour cet ouvrage a reçu une réponse phénoménale, qui s'est manifestée autant par le nombre de textes soumis que par leur qualité. Nous y voyons un signe que ce projet a été lancé au bon moment. Les travaux rejoignent les thématiques clés de l'Axe et contribuent à redéfinir nos propres champs d'intérêt et d'expertise : politiques culturelles, industries créatives, enjeux juridiques et de droit comparé, diversité des expressions culturelles, patrimoine, médias et journalisme. Plusieurs études sont de nature exploratoire ou portent sur des pratiques expérimentales, ce qui atteste du caractère embryonnaire du domaine. En ce sens, l'ouvrage constitue à la fois un guide pour la réflexion et un marqueur (*benchmark*) pour mesurer l'évolution de ces travaux dans les années à venir.

Finalement, ce sont 21 textes auxquels contribuent 39 auteurs et autrices du Québec et d'ailleurs – dont plusieurs sont de jeunes chercheurs – qui sont rassemblés dans cet ouvrage divisé en quatre parties.

PARTIE I – L'INTELLIGENCE ARTIFICIELLE ET LA CRÉATION ARTISTIQUE

La première partie de l'ouvrage porte sur les défis multiples posés par l'IA à l'acte de création. **François Levin** propose de reformuler la question du potentiel créatif des machines, malgré l'absence d'une intention artistique, au-delà de la richesse de la réception et du processus d'élaboration des œuvres. Il faudrait plutôt s'interroger sur ce qui permettrait de créer de l'inattendu, de la surprise, « l'irruption du dehors à l'intérieur d'un processus déterministe ». On retrouve des questionnements similaires dans plusieurs autres contributions à l'ouvrage.

La capacité à produire de la surprise – ou à tout le moins de la confusion – et les possibilités de cocréation sont explorées par **Christine Blais**

14. L'Axe Arts, médias et diversité culturelle organise notamment un séminaire annuel qui aborde divers enjeux relatifs aux technologies numériques et à l'IA. Pour plus de détails, voir Observatoire international sur les impacts sociétaux de l'IA et du numérique, « Axe Arts, médias et diversité culturelle », <<https://observatoire-ia.ulaval.ca/axe/arts-medias-et-diversite-culturelle/>>.

et **Olivier Blais** à partir de l'installation en réalité augmentée *Chomsky vs. Chomsky*, intelligence conversationnelle vocale produite à partir d'un vaste répertoire d'interventions de l'auteur de la grammaire générative – et critique de l'IA. **Madelena Gonzalez** et **Camille Habault** s'intéressent à l'IA au théâtre, autant comme thème que comme élément de mise en scène, dans une réflexion visant à dépasser les inquiétudes et l'opposition entre *technè* et acte créatif.

Les deux chapitres suivants se penchent sur la place de l'IA dans l'univers du jeu. **Samuel Francblu** explique comment l'automatisation peut servir à envisager différents scénarios lors de la conception des jeux de société analogiques, en appuyant son analyse sur une modélisation des structures de jeu, soit leurs usages et mises en pratique par les joueurs et les joueuses. **Laura Iseut Lafrance St-Martin** et **Maude Bonenfant** offrent un état des lieux et évoquent des enjeux soulevés par la conception automatisée des jeux vidéo, tout particulièrement la génération procédurale qui se nourrit des données produites par les usagers pour développer de nouvelles itérations d'un jeu.

PARTIE II – LE RECOURS À L'INTELLIGENCE ARTIFICIELLE ET AUX ALGORITHMES DE RECOMMANDATION POUR LA DIFFUSION DES CONTENUS JOURNALISTIQUES ET CULTURELS

La deuxième partie de l'ouvrage se penche sur les médias d'information et les plateformes de contenus culturels ainsi que leurs publics respectifs, dont ils cherchent à capter, et le plus souvent monnayer, l'attention dans l'environnement numérique. Les algorithmes de recommandation demeurent des outils puissants en ce sens, mais les auteurs explorent aussi les possibilités de contournement pour les usagers et les pratiques spécifiques d'IA chez les acteurs médiatiques.

Grâce à une série d'exemples d'expérimentations en IA dans des médias d'information nord-américains, **Nicolas Saint-Germain** et **Patrick White** font ressortir la diversité des applications potentielles, les enjeux éthiques pour le journalisme et les inégalités entre acteurs, autant quant à la capacité d'investissement qu'à l'accès à l'expertise.

Édouard Germain, **Justine Dorval** et **Éric George** montrent les limites des algorithmes de recommandation des plateformes de diffusion vidéo en continu, qu'elles soient transnationales et d'origine américaine ou conçues pour le marché canadien. Celles-ci, comme tout dispositif

technique, laissent toujours une place au contournement des règles d'usage prévues par les concepteurs. À partir d'une enquête qualitative auprès de jeunes adultes au Québec, **Lily-Cannelle Mathieu** a produit une liste de facteurs facilitant et freinant l'agentivité des usagers des plateformes. L'étude d'**Alix Bénistant**, **Christine Chevret-Castellani** et **Sarah Labelle** sur la construction des profils Netflix explore les tactiques d'usage et les capacités critiques des personnes face aux logiques algorithmiques.

L'enjeu de la découvrabilité, ou de la visibilité des produits culturels sur les plateformes numériques, est abordé de manière fort originale dans le chapitre suivant. **Ève Givois**, **Éric Kergosien**, **Laure Bolka-Tabary**, **Samuel Gantier** et **Bernard Jacquemin** ont étudié la plateforme coopérative Tënk, dédiée au cinéma documentaire d'auteur, structurée par un processus d'éditorialisation par des bénévoles experts, auquel se greffent la recommandation sociale et l'expérimentation d'une recommandation automatisée. Enfin, sur une note plus optimiste, **Viêt Cao** et **Juliette Denis** plaident pour des partenariats de données dans le domaine de la culture et fournissent de nombreuses pistes et de nombreux exemples pour leur mise en œuvre.

PARTIE III – L'INTELLIGENCE ARTIFICIELLE ET LA SAUVEGARDE DU PATRIMOINE CULTUREL

La troisième partie de l'ouvrage nous mène sur le terrain du patrimoine culturel, envisagé sous diverses formes, matérielles et immatérielles. Tout en soulignant les risques que les systèmes d'IA peuvent représenter dans ce domaine, ce sont plutôt ses potentialités qui sont mises en exergue.

Bien qu'ils évoquent d'emblée la fracture numérique persistante, **Serigne Momar Sarr** et **Alassane Koringho Sagna** s'intéressent aux répercussions de l'IA sur le patrimoine vivant de l'Afrique et à la manière dont elle peut être utilisée pour promouvoir les langues et stimuler la diffusion des savoirs. Les auteurs se penchent en particulier sur les évolutions en cours dans le domaine de l'IA au Sénégal et les solutions créées pour favoriser l'apprentissage et l'usage des langues locales. L'IA est ici perçue comme un outil apte à favoriser l'alphabétisation, tout en tenant compte des réalités socioculturelles.

Alors qu'il est devenu relativement commun de recourir à l'IA pour générer une version écrite de fichiers audio, son utilisation dans la retranscription automatique de textes manuscrits paraît inédite. **Emmanuelle Perrin**

et **Philippe Chassignet** en explorent le potentiel, plus précisément aux fins de retranscription d'archives archéologiques datant du XIX^e siècle, une démarche menée dans le cadre d'un projet participatif. Le projet « Bulliot, Bibracte et moi » incite ainsi à entrevoir l'IA comme une clé permettant de découvrir un patrimoine méconnu et difficilement accessible.

Cette clé que constitue l'IA est aussi à la portée du guide touristique, interprète du patrimoine culturel. En traitant de l'introduction de l'IA dans le métier de guide et en explorant ses répercussions sur la pratique de la visite guidée, **Lina Uzlyte** examine les pratiques nouvelles qui se développent dans le domaine touristique. Ses questions portent en outre sur la manière d'intégrer l'IA dans un métier dont l'exercice est encadré par des normes et des principes éthiques ; le guide assume en effet une responsabilité envers le public et, à ce titre, l'influence des systèmes d'IA sur l'activité de ce médiateur culturel mérite que l'on s'y attarde.

PARTIE IV – LES RÉPERCUSSIONS DE L'INTELLIGENCE ARTIFICIELLE SUR LA PROPRIÉTÉ INTELLECTUELLE, LES DROITS HUMAINS ET LA DIVERSITÉ CULTURELLE

La quatrième et dernière partie de l'ouvrage rassemble sept textes qui ont pour trait commun d'aborder divers enjeux et problèmes soulevés par l'IA et liés au respect des droits humains.

Trois premiers textes explorent sous divers angles la relation entre l'IA et les régimes de propriété intellectuelle. **Alexandra Bensamoun** évoque d'entrée de jeu le lien inextricable entre les données et l'IA – les premières étant indispensables au fonctionnement de la seconde – avant de plonger dans une réflexion sur l'applicabilité du droit d'auteur aux créations algorithmiques, c'est-à-dire aux œuvres générées par une IA. Elle examine notamment les conditions de la protection de telles œuvres, en s'interrogeant sur la possible qualification de « création de forme originale ». L'auteure aborde en outre la relation entre l'œuvre et les données, notamment l'œuvre en tant que donnée utilisée pour nourrir les systèmes d'IA qui, à leur tour, génèrent de nouvelles œuvres. Elle insiste sur le fait que, même appréhendée comme une donnée, l'œuvre demeure soumise au droit d'auteur.

Un autre regard sur la relation entre l'IA et la propriété intellectuelle est posé par **Camille Alegre**, selon laquelle il existerait une incompatibilité fondamentale entre l'une et l'autre, qui ne relève pas tant du domaine

juridique que du domaine terminologique. Selon elle, cette incompatibilité milite contre l'octroi de la qualité d'auteur à l'IA. Elle soutient qu'au fil de l'évolution et du développement de l'IA le nombre de parties prenantes humaines engagées dans le processus créatif augmente. Par conséquent, un processus créatif émancipé de tout travail humain pourrait n'être que fiction.

Enfin, le « piratage créatif » (*creative piracy*) est le sujet abordé par **Stephen Monteiro** qui s'intéresse ainsi aux techniques de masquage destinées à déformer un contenu de manière à déjouer les systèmes d'IA auxquels les plateformes ont recours pour détecter les œuvres reproduites ou diffusées en violation des règles de propriété intellectuelle. L'auteur introduit dans sa réflexion les thèmes de l'esthétique numérique, de la production culturelle vernaculaire et de l'utilisation équitable pour remettre en question l'effet du masque qui, au final, pourrait alternativement être vu comme générateur de nouvelles œuvres, pavant ainsi la voie à un usage dit loyal.

Le champ de la relation entre l'IA et les droits humains est investi plus largement par **Lucia Flores Echaiz** qui aborde la gouvernementalité algorithmique. Visant à cibler les besoins et les préférences des utilisateurs en vue de personnaliser les offres commerciales, la gouvernementalité algorithmique génère une bulle de filtre – nommée par l'auteur « bulle immunitaire hyper-personnalisée » – qui menace, selon elle, plusieurs droits fondamentaux, dont la liberté d'expression et d'opinion, le droit à la vie privée et le droit à l'égalité.

De manière plus implicite, l'IA et les droits humains constituent aussi la toile de fond des réflexions menées par **Iris Richer** et **Clémence Varin** qui abordent les enjeux relatifs à la préservation de la diversité culturelle. Les auteures explorent plus précisément la manière dont les préoccupations culturelles sont intégrées – ou non – dans les instruments relatifs à l'encadrement de l'IA et plaident pour le développement d'une IA culturellement responsable, c'est-à-dire respectueuse de la diversité culturelle.

C'est finalement sur le terrain de la discrimination que nous amènent les auteurs des deux derniers textes de cet ouvrage. Face à la propagation de normes discriminatoires envers les femmes dans les algorithmes, **Eleonore Fournier-Tombs** et **Céline Castets-Renard** se penchent sur les préjugés sexistes entretenus par la féminisation de l'IA. Prenant appui sur l'exemple des agents conversationnels et des systèmes d'assistance vocale, les auteures dénoncent l'omniprésence de l'IA féminisée qui encourage

des interactions sexuées stéréotypées et présentent une série de moyens permettant de réduire les biais systémiques envers les femmes.

La censure des contenus LGBT sur les médias sociaux, rendue possible par un recours à l'IA dans la modération, est enfin le cœur de la réflexion menée par **Thibault Grison** et **Virginie Julliard**. Les auteurs relatent que, face à de telles pratiques de discrimination algorithmique et d'invisibilisation de certains contenus, les communautés censurées se mobilisent afin d'établir des stratégies sémiotiques leur permettant de déjouer les mécanismes de modération automatisés et de se rendre de nouveau visibles.

* * *

Plus qu'un outil de transmission des connaissances sur la manière dont l'IA investit désormais les champs de la culture et des médias, cet ouvrage est aussi le miroir des perceptions ou des intuitions du moment. Il traduit en quelque sorte l'état d'esprit engendré par une technologie – l'intelligence *artificielle* – qui s'impose dans un domaine de la *création*, de contenus culturels ou médiatiques, jusqu'à récemment perçu comme indissociable du génie *humain*. Or, si plusieurs textes témoignent d'une certaine confiance placée dans l'IA pour accompagner le créateur, pour augmenter son potentiel, voire pour le protéger, une majorité d'auteurs partagent surtout leurs inquiétudes ou leurs interrogations. Il faudra suivre avec attention la manière dont ces préoccupations évolueront au fil du temps ; au fil de l'essor de l'IA dans tous les aspects de nos sociétés ; de son essor dans toutes les dimensions de notre vie culturelle.

PARTIE I

L'INTELLIGENCE ARTIFICIELLE ET LA CRÉATION ARTISTIQUE

1

SURPRISE, INTERPRÉTATION, SURGISSEMENT

PENSER LA NOUVEAUTÉ À L'AUNE DES MACHINES CRÉATIVES

François Levin¹

Un certain nombre d'algorithmes d'intelligence artificielle contemporains, particulièrement d'apprentissage machine, sont considérés comme « créatifs » ou, du moins, ont des effets que l'on pourrait apparenter à de la création. Les exemples sont connus et certains ont été amplement commentés récemment dans les médias et ont ainsi acquis un statut emblématique sur cette question : c'est le cas du *Portrait d'Edmond de Bellamy*, par exemple, dont la production a été « pilotée », pourrait-on dire, par le collectif Obvious, mais également des œuvres produites par Ai-Da², une intelligence artificielle humanoïde dont le « travail » a été exposé à Oxford en 2019, ou encore des productions d'AICAN³, qui ont également été exposées. Certains et certaines artistes utilisent les capacités « créatives » des algorithmes de manière plus structurée et approfondie. C'est le cas de Memo Akten⁴ ou encore de Refik Anadol qui utilisent les logiciels de reconnaissance visuelle pour produire de nouvelles formes. La

1. Doctorant en philosophie à l'École polytechnique.

2. Ai-Da est un robot humanoïde créé par une société britannique ; elle a peint un certain nombre de tableaux, qui ont fait l'objet d'une exposition en Angleterre en 2019.

3. AICAN est un algorithme d'un genre nouveau, un *creativ adversarial network*. Voir AICAN. *Art of the Future, Now*, <<https://aican.io/>>.

4. Voir notamment Deep Meditations ou Deep dream is blowing my mind.

série *Unsupervised — Machine Hallucinations*⁵ de ce dernier est particulièrement frappante à cet égard. Dans le domaine littéraire⁶ et musical⁷ et scientifique⁸, ces algorithmes sont également très utilisés. Ces évolutions algorithmiques ont occasionné un renouveau des réflexions autour de la créativité computationnelle et de l'imagination artificielle⁹, qui s'inscrit dans un contexte plus large de réflexions nouvelles autour de la notion de créativité, notamment en psychologie.

En parallèle, de nouveaux modèles d'apprentissage profond appliqués au traitement du langage naturel ont été développés, tels que GPT-3¹⁰ d'Open AI, et ont augmenté énormément l'efficacité de la génération automatique de textes, ouvrant davantage la voie à l'écriture de fictions, de scénarios et de textes théoriques par des algorithmes¹¹. Il est possible de tracer rapidement, de manière liminaire, une typologie des processus qui mènent aux résultats que l'on considère comme « créatifs » :

- La corrélation inductive : certains algorithmes d'apprentissage supervisé ou non supervisé permettent de déterminer des corrélations jusqu'alors non découvertes, et peuvent ainsi conduire à produire des résultats qui pourraient être qualifiés de créatifs si un chercheur y parvenait. C'est le cas pour les algorithmes de fouille automatisée de textes et de données (*text and data mining*), qui peuvent conduire à déterminer des corrélations entre différents éléments qui étaient jusqu'alors tenus éloignés. Dans le domaine de la recherche, les algorithmes de fouille de textes,

5. Voir <<https://refikanadol.com/works/unsupervised/>>.

6. Le livre *1 the Road*, composé par une intelligence artificielle, a récemment fait l'objet d'une couverture médiatique importante.

7. Les morceaux d'ambiance pour des publicités, le cinéma et les jeux vidéo sont de plus en plus composés par des intelligences artificielles. Voir cet article pour plus de détails : *La lettre du musicien*, « L'intelligence artificielle : la musique du futur », <https://www.lalettredumusicien.fr/s/articles/6398_0_lintelligence-artificielle-la-musique-du-futur>.

8. Le programme AlphaFold est récemment parvenu à déterminer la forme d'une protéine à partir d'une séquence d'acides aminés déterminée. Voir Anouk Tomas, « AlphaFold, l'intelligence artificielle révolutionnaire pour la recherche en protéomique », *Sciences et Avenir*, 2 décembre 2020, <https://www.sciencesetavenir.fr/fondamental/biologie-cellulaire/alphafold-l-intelligence-artificielle-revolutionnaire-pour-la-recherche-en-proteomique_149674>.

9. Un séminaire tenu à l'école normale supérieure de Paris sur ce thème en est notamment témoin : *Fiction de l'artificiel postdigital*, <<http://postdigital.ens.fr/>>.

10. GPT-3 est un modèle de langage développé par la société OpenAI annoncé le 28 mai 2020 et ouvert aux utilisateurs par l'API d'OpenAI en juillet 2020. Au moment de son annonce, GPT-3 était le plus gros modèle de langage jamais formé avec 175 milliards de paramètres.

11. Sur cette question, voir le numéro spécial de Justin Weinberg, « Philosophers on GPT-3 », *Daily Nous*, 30 juillet 2020, <<https://dailynous.com/2020/07/30/philosophers-gpt-3/>>.

malgré un certain nombre d'obstacles liés à la propriété intellectuelle notamment¹², ont pu faire la preuve de leur efficacité.

- La génération aléatoire contrôlée : on peut ainsi désigner le fonctionnement d'algorithmes spécifiques, les réseaux antagonistes génératifs (*generative adversarial networks*, ou GAN), ainsi que les réseaux antagonistes créatifs (*creative adversarial networks*, ou CAN). Ces algorithmes, comme le nom du dernier l'indique, ont été à l'origine des résultats les plus spectaculaires en matière de créativité et c'est souvent à leur propos que l'on parle d'intelligence artificielle créative. Il s'agit d'algorithmes composés de deux parties : un algorithme générateur, qui produit des images, par exemple, et un algorithme discriminateur, issu d'un apprentissage supervisé, qui tente de déterminer si ce que produit le générateur correspond au résultat pour lequel il a été entraîné. À ce fonctionnement, qui est celui des GAN, les inventeurs des CAN ont ajouté un principe de déviation par rapport aux styles établis, afin d'optimiser leur potentiel créatif (Elgammal, Liu, Elhoseiny et Mazzone, 2017). Ainsi peuvent être produites des œuvres plastiques ou littéraires, par exemple, notamment en reproduisant le style d'un artiste. La plupart des générateurs de textes fonctionnent comme cela, ainsi qu'un certain nombre de dispositifs utilisés dans l'industrie pharmaceutique pour découvrir de nouvelles molécules¹³. Ces dispositifs fonctionnent en fait suivant un mécanisme de production aléatoire contrôlée, qui peut se rapprocher, d'une manière lointaine, de l'abduction décrite par Charles S. Peirce, c'est-à-dire de la formulation d'hypothèses informées afin de résoudre un problème.
- L'éloignement par rapport à la norme : certains algorithmes ont un mode d'apprentissage précis, que l'on désigne sous le terme d'« apprentissage par renforcement ». Dans ce type d'apprentissage, l'algorithme apprend à partir de son interaction avec l'environnement, sans que lui soit fourni en amont un

12. Voir notamment pour comprendre le problème au niveau européen : Marc Rees, « Le chantier de l'exception "text and data mining" au ministère de la Culture », *Nextinpact*, 31 août 2020, <<https://www.nextinpact.com/article/43410/le-chantier-lexception-text-and-data-mining-au-ministere-culture>>.

13. Un exemple de ces dispositifs : le discriminateur a appris, à partir d'une base de données, les caractéristiques qui correspondent à l'objectif final du médicament, c'est-à-dire les propriétés chimiques nécessaires pour les molécules ; le générateur produit des molécules qui correspondent à ce modèle.

ensemble de données sur lesquelles s'entraîner. À la suite de son interaction avec l'environnement (par exemple, un coup sur un plateau de jeu), une récompense lui est attribuée ou non, afin qu'il parvienne à améliorer son comportement. Ont pu ainsi être développés des algorithmes dans le domaine du jeu, dont AlphaGo évidemment, qui a fait couler beaucoup d'encre, mais aussi, plus récemment, AlphaStar, qui joue à Starcraft 2, un jeu de stratégie en ligne¹⁴, dont le nombre de possibilités est encore plus élevé que le jeu de go. Ce type d'algorithme produit des « coups » qui sont très surprenants pour ceux et celles qui sont habitués à jouer au jeu en question et qui paraissent dénoter une forme d'écart par rapport à la norme attendue du jeu que l'on peut qualifier comme une forme de créativité : cela a été notamment le cas avec le fameux coup n° 37 d'AlphaGo contre Lee Sedol, le champion du monde en titre, jugé comme une erreur par beaucoup de commentateurs avant que la victoire finale d'Alpha Go les amène à réviser ce jugement.

- L'erreur ou l'accident : certaines formes de ce qu'on pourrait reconnaître comme des résultats créatifs produits par l'intelligence artificielle relèvent d'un accident ou d'une erreur. Ainsi, les erreurs dans la reconnaissance de formes par les algorithmes de reconnaissance visuelle donnent lieu à des images que l'on a pu qualifier de « rêves algorithmiques » et qui ont pu être décrites comme faisant preuve d'une forme d'imagination artificielle¹⁵.

1. LE PROBLÈME GÉNÉRAL : LA MACHINE CRÉATIVE

La question que nous allons nous poser dans cet article est la suivante : comment interpréter ces capacités nouvelles ? Comment rendre compte théoriquement du trouble qu'elles suscitent, en brouillant les catégories installées entre, d'un côté, la création et, de l'autre le fonctionnement machinique ? Une machine peut-elle vraiment créer, c'est-à-dire produire quelque chose de nouveau et ayant une valeur ou ne peut-elle qu'imiter la création ? En effet, alors que la création de nouveautés semble relever de

14. Pour en savoir plus sur Starcraft2, voir : Julien Lausson, « Pour battre l'IA de Google sur StarCraft 2, vous devez être dans les 0,2 % des meilleurs joueurs du monde », *Numerama*, 31 octobre 2019, <<https://www.numerama.com/sciences/565976-vous-ne-pourrez-plus-jamais-battre-lia-de-google-sur-starcraft-2-seuls-les-02-meilleurs-joueurs-resistent.html>>.

15. Notamment par l'artiste Gregory Chatonski, « Terre Seconde – Grégory Chatonsky et Emanuele Coccia – Audi Talents », <<https://www.youtube.com/watch?v=90EPfupUsAs>>.

l'imprévisible, de l'irruption, voire de l'inspiration, comment imaginer qu'une machine, c'est-à-dire un système déterministe, puisse en être à l'origine ? La « machine créative » n'est-elle qu'une contradiction dans les termes et la création algorithmique un abus de langage ?

L'objectif de cet article n'est pas de répondre directement à cette question, mais de recenser quelques grandes manières possibles de la poser et, par là, même d'y répondre. La question « une machine peut-elle vraiment créer ? » connaît plusieurs formulations, que nous allons passer en revue : 1) une machine peut-elle produire quelque chose de surprenant ? ; 2) une machine peut-elle produire quelque chose qui puisse faire l'objet d'une interprétation ? ; 3) une machine reproduit-elle les processus spécifiques (combinaison, exploration, transformation) qui caractérisent le fait de créer ? ; 4) une machine intègre-t-elle dans son processus une « altérité » telle que les effets de son fonctionnement ne soient pas entièrement contenus dans les causes qui le déterminent ?

On peut dire que les deux premières questions relèvent d'une position « relationnelle » quant à la question de la création, c'est-à-dire d'une position qui définit la création à partir de la relation qu'entretiennent les individus envers l'objet censé être une création, tandis que les deux dernières relèvent d'une position « processuelle », c'est-à-dire qui définit la création comme un processus d'un genre particulier.

2. LA POSITION RELATIONNELLE

2.1 Retour sur l'argument de Lady Lovelace

Le problème de la machine créative n'est pas tout à fait nouveau et il a déjà reçu une formulation demeurée célèbre sous la plume d'un des devanciers les plus fameux de l'intelligence artificielle moderne, Alan Turing. Dans son article de 1950 (Turing, 1950), dans lequel il formule pour la première fois l'idée de son test, celui-ci formule en effet le problème de la machine créative sous la forme de ce qu'il appelle l'argument de Lady Lovelace. Lady Ada Lovelace était une aristocrate anglaise du 19^e siècle, considérée comme la première programmeuse de l'histoire du fait de son mémoire sur la machine à calculer de Charles Babbage. Turing la cite ainsi dans son article :

Our most detailed information of Babbage's Analytical Engine comes from a memoir by Lady Lovelace (1842). In it she states, « The Analytical Engine has no

pretensions to originate anything. It can do whatever we know how to order it to perform » (*her italics*)¹⁶ (Turing, 1950).

Turing cite cet argument comme un des principaux contre l'idée qu'une machine puisse penser dans un article qui essaie précisément de montrer les conditions de possibilité d'une telle affirmation, et qui formule à cet égard le « test » qui est resté célèbre sous le nom de « test de Turing¹⁷ ».

Turing fait ici référence à ce qu'écrit Ada Lovelace dans les notes de sa traduction anglaise du mémoire de L.F. Menabrea concernant la machine analytique de Charles Babbage. Ces notes dépassent en volume et en intérêt le mémoire lui-même. Dans la note G, la plus célèbre, puisqu'elle y développe ce que l'on peut considérer comme le premier programme informatique, destiné à calculer les nombres de Bernoulli grâce à la machine de Babbage, elle affirme l'incapacité, pour cette machine, de formuler quelque chose de nouveau. Par parenthèse, on pourrait, avant d'entrer dans l'analyse de l'argument précis de Turing, tenter de le nuancer quelque peu, en revenant rapidement au texte originel de Lady Lovelace, qui n'est trop souvent cité que par l'intermédiaire de Turing, qui n'en avait d'ailleurs eu lui-même connaissance que par les textes de Douglas Hartree, un mathématicien anglais du 20^e siècle : en effet, Lovelace formule juste avant l'extrait que nous avons cité une précaution méthodologique, qui pourrait d'ailleurs servir de *vade-mecum* pour toute pensée de la technologie :

*In considering any new subject, there is frequently a tendency, first, to overrate what we find to be already interesting or remarkable ; and, secondly, by a sort of natural reaction, to undervalue the true state of the case, when we do discover that our notions have surpassed those that were really tenable*¹⁸ (Lovelace, 1842).

-
16. « Les informations les plus détaillées sur le moteur analytique de Babbage proviennent d'un mémoire de Lady Lovelace (1842). Elle y déclare : "Le moteur analytique n'a pas la prétention d'être à l'origine de quoi que ce soit. Il peut faire *tout ce que nous savons lui ordonner de faire*" (italiques de l'auteure) » (notre traduction).
 17. Pour une présentation précise de ce test et du jeu des arguments et des contre-arguments qu'il a suscités, voir Matthieu Tricot, *Le moment cybernétique : la constitution de la notion d'information*, Seyssel, Champ Vallon, 2008.
 18. « En considérant tout nouveau sujet, on a souvent tendance, d'une part, à *surestimer* ce qu'on trouve d'emblée intéressant ou remarquable ; et, d'autre part, par une sorte de réaction naturelle, à *sous-estimer* le véritable état de la question, lorsque nous constatons que nos notions ont dépassé celles qui étaient réellement acceptables » (notre traduction).

Après cet avertissement, Lady Lovelace exprime une idée dont Turing ne rend pas compte dans son article :

But it (the analytic Machine) is likely to exert an indirect influence and reciprocal influence on science itself in another manner. For, in so distributing and combining the truth and the formula of analysis, that they may become most easily and rapidly amenable to the mechanical combinations of the engine, the relations and the nature of many subjects in that science are necessarily thrown into new lights, and more profoundly investigated. [...] [I]n devising for mathematical truths a new form in which to record and throw themselves out for actual use, views are likely to be induced, which should again react on the more theoretical phase of the subject¹⁹ (Lovelace, 1842).

Comment comprendre cette affirmation selon laquelle le fait d'organiser les éléments d'une démonstration de manière à ce qu'elle soit calculable par la machine de Babbage – ce que fait d'ailleurs Lovelace à la suite de ce développement avec les nombres de Bernoulli – aurait une influence indirecte et seconde sur la manière de concevoir le problème en question ? Cela semble signifier que la disposition nouvelle des arguments et leur séquençage permettent de concevoir de nouvelles pistes démonstratives, voire de nouveaux problèmes, ne serait-ce que par l'effet visuel que produit cette disposition. La « mise en algorithme » aurait ainsi une valeur heuristique du fait de la présentation nouvelle des éléments dont elle serait à l'origine. Si cette intuition n'est pas un argument en faveur d'une machine créative, qui d'elle-même pourrait produire du nouveau, il est tout de même intéressant de souligner qu'elle permet de nuancer la vision thuringienne de « l'argument de Lady Lovelace ». Il est d'ailleurs intéressant de noter que le raisonnement de Lady Lovelace sur l'effet heuristique indirect de la « mise en algorithme » d'un problème mathématique peut trouver un écho dans la démarche démonstrative de Turing lui-même, dans l'article de 1936 cette fois, sur la fameuse « machine de Turing » (Turing, 1936). C'est en effet en produisant l'idée d'une machine démonstrative et de la mécanisation totale

19. « Mais elle est susceptible d'exercer une influence indirecte et une influence réciproque sur la science elle-même d'une autre manière. En effet, en distribuant et en combinant la vérité et la formule de l'analyse de manière à ce qu'elles se prêtent le plus facilement et le plus rapidement possible aux combinaisons mécaniques de la machine, les relations et la nature de nombreux sujets de cette science sont nécessairement éclairées d'un jour nouveau et étudiées de manière plus approfondie. C'est une conséquence décidément indirecte et quelque peu spéculative d'une telle invention. Il est cependant assez évident, de manière générale, qu'en concevant pour les vérités mathématiques une forme nouvelle pour leur développement, il est possible que de nouvelles visions en surgissent, qui devraient avoir des répercussions sur la phase plus théorique du sujet » (notre traduction).

du calcul, dont Babbage était un des initiateurs théoriques, que Turing parvient à répondre à la question de la décidabilité, posée par Hilbert dans son programme de refondation des mathématiques. Ainsi, c'est en passant par la « mise en algorithme » du calcul que Turing parvient à produire une « nouvelle vision », pour reprendre les mots de Lovelace, concernant le problème fondationnel posé par Hilbert. Il produit en effet l'idée, et la démonstration, du caractère indécidable des systèmes formels, ruinant ainsi l'édifice hilbertien, quelques années après les premiers coups de boutoir portés par Gödel, relativement à la question de la complétude et de la consistance. Il est donc remarquable, par parenthèse, qu'Ada Lovelace paraisse formuler, dans son intuition de l'effet indirect des algorithmes sur les évolutions de la théorie scientifique, le type de raisonnement même qui sera celui de Turing dans son article de 1936.

2.2 Une machine peut-elle produire quelque chose de surprenant ?

La réponse que formule Turing à l'argument de Lady Lovelace s'appuie sur une reformulation de cet argument dans l'article de 1950 : il s'agit de la première reformulation de la question « une machine peut-elle créer ? » en « une machine peut-elle produire quelque chose de surprenant ? »

A better variant of the objection says that a machine can never « take us by surprise ». This statement is a more direct challenge and can be met directly. Machines take me by surprise with great frequency. This is largely because I do not do sufficient calculation to decide what to expect them to do, or rather because, although I do a calculation, I do it in a hurried, slipshod fashion, taking risks. Perhaps I say to myself, « I suppose the voltage here ought to be the same as there : anyway let's assume it is. Naturally I am often wrong, and the result is a surprise for me for by the time the experiment is done these assumptions have been forgotten »²⁰ (Turing, 1950).

Cette reformulation par Turing est en fait une manière tout à fait différente de poser la question, puisqu'il s'agit de produire une équivalence entre la notion de nouveauté et la notion de surprise. Turing affirme

20. « Une meilleure variante de l'objection dit qu'une machine ne peut jamais “ nous prendre par surprise ”. Cette déclaration est plus claire et peut être discutée directement. Les machines me surprennent très fréquemment. C'est en grande partie parce que je ne fais pas suffisamment de calculs pour savoir à quoi m'attendre, ou plutôt parce que, bien que je fasse un calcul, je le fais de façon précipitée, bâclée, en prenant des risques. Peut-être que je me dis : “ Je suppose que la tension électrique ici devrait être la même que là-bas ; de toute façon, supposons qu'elle le soit. Naturellement, je me trompe souvent, et le résultat est une surprise pour moi car, au moment où l'expérience est faite, j'ai oublié les hypothèses que j'avais formulées » (notre traduction).

qu'une machine peut nous prendre par surprise et que c'est d'ailleurs très souvent le cas. Cette surprise provient, selon ses dires, de l'insuffisance des calculs qu'il a effectués et qui ne lui permettent donc pas de prévoir parfaitement ce que va produire la machine. C'est donc une définition de la nouveauté que l'on pourrait appeler relationnelle, construite depuis la réception, c'est-à-dire à partir de la relation qu'entretient une personne avec le résultat du fonctionnement d'une machine ou d'un algorithme. Cette affirmation selon laquelle les machines peuvent nous surprendre semble pouvoir s'appliquer tout à fait aux algorithmes contemporains, qui, du fait de leur complexité extrême et du nombre des paramètres qu'ils mettent en jeu dans le cadre de l'apprentissage profond, rendent très complexe la prévision de leurs effets. Ainsi, alors que Turing trouvait dans le calcul erroné ou bâclé la source de cette surprise potentielle, on pourrait affirmer que les machines contemporaines sont de fait des machines surprenantes, tant leur complexité est extrême, ce qui rend les calculs pour le moins compliqués, sinon impossibles. C'est d'ailleurs tout l'enjeu de la réflexion éthique autour des « boîtes noires²¹ », que de trouver des manières de savoir ce que la machine va effectuer et de ne pas être, sinon surpris, du moins désemparé par ses résultats. Ainsi, suivant cette première définition, une machine peut produire du nouveau, c'est-à-dire des résultats surprenants.

2.3 Une machine peut-elle produire quelque chose qui peut faire l'objet d'une interprétation ?

La deuxième manière de penser la production du nouveau à partir du point de vue de la réception et non du processus lui-même consiste à faire de l'interprétation le lieu de constitution de cette nouveauté. L'interprétation du spectateur serait en vérité productrice de l'œuvre, ou du moins de son caractère d'œuvre, sans même que soit postulée une intention qui préside à l'œuvre. L'interprétation joue en effet une place centrale dans la constitution de certaines œuvres : c'est déjà le cas, sans même évoquer l'intelligence artificielle, pour les systèmes de règles appliquées à la production littéraire. Les cent mille milliards de poèmes²² de Raymond Queneau sont ainsi un

21. Les algorithmes d'apprentissage machine sont très complexes et il est souvent très difficile de suivre le parcours qui a mené à ce qu'une décision soit prise plutôt qu'une autre. Ils ont donc pu être qualifiés de « boîtes noires » et une réflexion éthique s'est développée autour de la meilleure manière de renforcer la transparence de ces systèmes. Voir notamment Frank Pasquale, *The Black Box Society : les algorithmes secrets qui contrôlent l'économie et l'information*, Limoges, FYP Éditions, 2015.

22. Raymond Queneau, *Cent mille milliards de poèmes*, Paris, Gallimard, 1961. Cet ouvrage est composé de dix sonnets superposés suivant un système permettant de combiner tous les

exemple du fait qu'une « machine littéraire » peut produire de véritables œuvres à partir d'un jeu combinatoire et syntaxique, notamment parce que l'interprétation du lecteur participe en grande partie à produire la signification. Cette interprétation peut donc se passer entièrement d'un postulat d'intention. Cette conception d'une œuvre comme un pur objet d'interprétation sans avoir besoin de recourir à une intention de l'auteur ne s'applique pas uniquement aux œuvres produites par des systèmes de règles formelles, mais potentiellement à toute œuvre. Ainsi l'analyse structuraliste des œuvres d'art, notamment des textes littéraires, les conçoit comme des systèmes de signes, objets potentiels d'une pure analyse formelle et de l'application de procédés interprétatifs déliés entièrement de l'idée d'une intention de l'auteur. C'est le cas notamment dans la célèbre analyse du poème *Les chats* de Baudelaire par Jakobson et Lévi-Strauss (Jakobson et Lévi-Strauss, 1962). Suivant cette conception, on peut imaginer qu'une machine puisse créer une pure œuvre combinatoire, sans intention, qui soit pourtant l'objet d'une analyse formelle et d'interprétations à l'instar d'une œuvre de Charles Baudelaire et dont la signification soit entièrement construite depuis la réception.

L'absence évidente d'intention de la machine ne supprimerait pas le caractère créatif de l'algorithme, qui est construit entièrement depuis le lieu de la réception. C'est d'ailleurs ce qu'affirme André Ourednik dans *Robopoïèses* : « Nous savons ainsi déjà que l'interprète détermine une part essentielle du sens d'une œuvre. Mais la machine à écrire intelligente nous montre aujourd'hui qu'il peut le déterminer entièrement. [...]. L'auteur au creux de la machine, le *Ghost in the shell*, naît de nos émotions et de nos imaginaires » (Ourednik, 2021, p. 146). Ainsi la définition de la nouveauté par la réception recouvre un certain nombre de formes différentes, qui engagent différemment la subjectivité de celui ou celle qui est le sujet de la réception. Dans la surprise, la réception implique faiblement la subjectivité, contrairement à ce qu'il se passe dans l'œuvre, qui est construite par l'interprétation. Davantage même que le sens, l'interprète participe à la création de l'œuvre robotique, à la robopoïèse : « Notre interprétation est la part *impensée* de leur projet ; elle libère le robot créateur de son asservissement aux raisonnements de ses concepteurs. Nous jouons un rôle essentiel dans la *robopoïèse* ! » (Ourednik, 2021, p. 147).

On comprend ainsi que cette première manière de répondre à la question de la possibilité pour une machine de produire du nouveau consiste en un efficace évitement méthodologique : il ne s'agit pas de s'interroger directement sur la question du déterminisme, que pose en fait Lady Lovelace, mais de partir uniquement du rapport entre le résultat du processus et les individus qui en sont les destinataires.

3. UNE DÉFINITION DE LA NOUVEAUTÉ INSUFFISANTE

Cette définition de la nouveauté par la réception semble néanmoins insuffisante, pour différentes raisons. Elle ne permet pas, tout d'abord, de produire une compréhension de ce qu'est la nouveauté qui dépasserait la simple appréciation individuelle. C'est particulièrement vrai pour ce que Turing dit de la surprise. En effet, si l'on suit son raisonnement, il semblerait que le degré de nouveauté du résultat d'un système déterministe de calcul puisse varier en fonction du niveau en mathématiques de la personne qui effectue les calculs : plus les calculs qu'elle effectue contiennent d'erreurs, plus sa surprise à la découverte du résultat exact sera en effet importante. De la même manière, plus un système technique est compliqué, plus sa capacité à produire des résultats surprenants est importante et, partant, plus sa capacité à produire de la nouveauté serait également importante. Les systèmes contemporains d'apprentissage profond, dont les résultats peuvent être très surprenants pour un non-initié, et même, comme nous l'évoquions plus haut, pour les initiés eux-mêmes, seraient ainsi, suivant Turing, fortement producteurs de nouveauté. On comprend combien cette définition est insuffisante, tant elle ne permet pas de discriminer entre la nouveauté et la surprise, entre la création et son imitation. Si l'inattendu peut être une forme subjective du nouveau, construire une équivalence entre le surprenant et le nouveau ne peut conduire qu'à des affirmations absurdes. La capacité d'un individu à être surpris par un processus ne peut être considérée comme équivalente à la capacité du processus à produire du nouveau. La surprise, pour un individu, apparaît plutôt comme la résultante d'un écart par rapport à une attente, ce dont Turing rend compte avec l'exemple du calcul bâclé. Dans ce cas, et peut-être même dans tous les cas de surprise, celle-ci ne consiste pas en un critère de nouveauté, mais en un critère d'ignorance. Ignorance, de la part de l'individu surpris, de la suite logique et parfaitement prévisible des événements qui conduisent au résultat surprenant. Nous ne sommes pas surpris par le nouveau, nous sommes surpris par un fait qui se démarque par rapport à notre horizon

d'attente. Ce que désigne la surprise, c'est l'inadéquation de notre horizon d'attente avec la structure causale des faits. Ainsi, le surprenant est ce qui est nouveau pour nous quand, en vérité, ce ne l'est précisément pas en soi. La preuve en est qu'à partir du moment où l'on comprend ce qui est à l'origine de ce qui nous a surpris, ce qui était surprenant ne l'est plus ; la surprise est passée. Le surprenant est donc le nom d'un nouveau appelé à ne plus l'être, c'est-à-dire à être réinscrit dans la suite causale qui le détermine. Le surprenant désigne le nouveau déceptif, son absence réelle dans le dévoilement d'un effet saisissant ; il est la structure d'un nouveau qui n'est qu'apparent. Le raisonnement de Turing semble donc paradoxalement devoir nous conduire à conclure l'inverse de ce qu'il était censé démontrer : le fait que les machines nous surprennent tendrait à prouver qu'elles ne peuvent précisément pas produire du nouveau.

Quant à l'idée du rôle prééminent que jouerait l'interprétation dans la constitution du caractère créatif d'une production, elle nous paraît également insuffisante.

En effet, l'interprétation s'applique à un objet qui est déjà présenté à l'individu comme méritant, précisément, une interprétation. C'est-à-dire que l'interprétation ne constitue pas un critère de démarcation entre ce qui est et ce qui n'est pas une œuvre. L'idée que le sens d'une œuvre est constitué par l'interprétation ne permet pas de déterminer ce qui mérite que s'y applique l'interprétation. Dans l'exemple des poèmes de Queneau et *a fortiori* dans celui des *chats* de Baudelaire, l'objet de l'interprétation est déjà constitué comme un objet méritant d'être interprété avant que celui-ci fasse l'objet d'une analyse. La filiation que crée André Ourednik entre l'idée selon laquelle la réception est partie prenante de la constitution du sens d'une œuvre et le fait qu'elle pourrait ainsi devenir l'instrument de constitution d'un artéfact en œuvre passe donc sous silence le fait que les artéfacts soumis à interprétation ont déjà fait l'objet d'une sélection, ou plutôt d'une élection, qui les qualifie en tant qu'œuvres. Même dans le cas des cent mille milliards de poèmes, c'est le fait que Raymond Queneau les ait écrits qui les qualifie à être interprétés ; le fait même qu'il ait renoncé en partie à l'idée de l'intention de l'auteur au profit d'une forme de « machine littéraire » est lui-même compris comme une intention de sa part et comme précisément ce qui doit faire l'objet d'une interprétation. Est-ce à dire qu'il n'est pas possible de se passer de l'idée d'intention et qu'il n'est donc pas possible de comprendre la valeur et la nouveauté d'un artéfact à partir de la simple réception ? Cela pose plus généralement la question de savoir s'il est possible de déterminer ce qu'est une œuvre d'art sans définir des

caractéristiques qui soient propres à l'œuvre elle-même. Peut-on imaginer définir une œuvre uniquement à partir du point de vue de la réception sans tomber pour autant dans un pur relativisme qui consisterait à faire de chacun le dépositaire d'un critère de l'œuvre d'art qui conduirait *in fine* à abolir la distinction entre œuvre et non-œuvre ?

Le philosophe George Dickie a tenté de répondre à ce problème dans *Art and the Aesthetic* (Dickie, 1974, p. 74). Cet ouvrage classique de la philosophie analytique de l'art a pour principe de décaler l'analyse de ce qu'est une œuvre d'art. Pour Dickie, les critères de détermination de ce qu'est une œuvre ne sont pas à chercher du côté des caractéristiques propres des artefacts, mais plutôt du côté de la réception ; mais cette réception est cette fois collective et déterminée par des règles qui constituent ce qu'il appelle « le monde de l'art²³ ». Dans ce cadre-là, c'est moins l'interprétation individuelle qui produit l'œuvre en tant qu'œuvre qu'un ensemble de conventions qui obéissent à des normes sociales. Ou plutôt pourrait-on dire qu'il existe une double constitution de l'œuvre en tant qu'œuvre et que la réception agit à deux moments différents. Une première constitution porte sur sa catégorisation en tant que « candidat à l'appréciation » :

Une œuvre d'art au sens classificatoire est (1) un artefact (2) tel qu'un ensemble de ses aspects fait que le statut de candidat à l'appréciation lui a été conféré par une personne ou un ensemble de personnes agissant au nom d'une certaine institution sociale (le monde de l'art) (Dickie, 1974, p. 34).

Ce n'est qu'après cette première constitution d'un artefact comme œuvre d'art qu'une deuxième constitution peut alors intervenir et faire de l'interprétation par le spectateur un lieu de co-fabrication de l'œuvre, au sens où l'entend Ourednik.

Dans le cas de ce qui est produit par les machines, il est tout à fait possible que le monde de l'art l'érige en tant qu'œuvre d'art. C'est d'ailleurs déjà arrivé puisqu'en 2018 le *Portrait d'Edmond de Belamy* a été vendu aux enchères pour plus de 400 000 dollars américains. De la même manière, comme nous l'avons évoqué plus haut, les œuvres déjà citées d'AICAN ont fait l'objet d'une exposition dans une galerie. Enfin une exposition a eu lieu en 2019 à Oxford pour exposer les œuvres d'Ai-Da (référence à Ada Lovelace), une intelligence

23. L'idée fondamentale de l'approche institutionnelle de l'art est que, pour expliquer ce qui fait qu'un objet donné est ou non une œuvre d'art, il faut s'intéresser non seulement aux propriétés intrinsèques de cet objet, mais aussi, et surtout, à la place qu'il occupe au sein du contexte institutionnel fourni par ce que l'on peut appeler « le monde de l'art ».

artificielle humanoïde. De la même manière que pour la *Fontaine* de Duchamp, le monde de l'art constitue un artefact en œuvre d'art. En ce sens, on pourrait affirmer que certaines productions algorithmiques sont des œuvres, au sens où elles sont reconnues comme telles.

Cette réponse nous laisse néanmoins un sentiment d'insatisfaction, pour deux raisons : tout d'abord, elle ne semble pas vraiment répondre au problème de la machine créative que nous avons posé initialement ; ou du moins sa manière d'y répondre consiste à éviter la question, à ne pas l'affronter en direct. Il ne s'agit plus de savoir si une machine peut vraiment produire quelque chose de nouveau, malgré le fait qu'elle suit un processus déterministe ou ne peut qu'imiter la création, mais il s'agit de s'intéresser uniquement aux conditions sociales de réception et de détermination de ce qu'est une œuvre d'art. En suivant le fil de la position subjective, nous avons perdu celui de la réflexion sur le déterminisme. Et, au fond, la question revient : si, de fait, des productions algorithmiques peuvent être reconnues comme des œuvres, c'est-à-dire comme des artefacts créatifs suivant le sens du monde de l'art, peut-on pour autant dire qu'il s'agit véritablement d'une production du nouveau, d'une création ? Sur ce point, la position purement conventionnaliste est insatisfaisante.

Par ailleurs, et c'est la deuxième raison que l'on peut évoquer pour justifier l'insatisfaction que nous procure cette position, on peut estimer que, du point de vue de la philosophie de l'art elle-même, la position conventionnaliste doit être dépassée également. On a ainsi assisté, notamment avec l'œuvre d'Arthur Danto (Danto, 2019), à un retour, dans la philosophie analytique de l'art, de la question des qualités de l'œuvre, en réponse aux impasses de la position conventionnaliste. Pour Danto, il est en effet tout de même nécessaire de postuler qu'il y a des caractéristiques particulières qui font qu'une œuvre est une œuvre et qu'elle est reconnue comme telle par le monde de l'art et est un objet d'interprétation. Ces caractéristiques ne sont pas matérielles ou perceptuelles, puisque, depuis le *ready-made*, tout objet peut devenir une œuvre : la particularité d'une œuvre d'art tient en effet à ce qu'elle possède une structure intentionnelle, qu'elle est à propos de quelque chose. De premier abord, une œuvre créée par une machine ne peut donc, à moins de rejouer le jeu de l'intention en le prêtant aux concepteurs de la machine eux-mêmes, être considérée comme une œuvre d'art, puisqu'elle ne possède pas de structure intentionnelle.

4. LA POSITION PROCESSUELLE

Il semble donc que la manière que l'on a appelée « relationnelle » de poser la question de la machine créative soit une impasse et qu'il faille plutôt se pencher sur le processus lui-même pour tenter de répondre à la question de savoir si une machine peut produire du nouveau.

4.1 Une machine peut-elle reproduire les processus spécifiques (combinaison, exploration, transformation) qui caractérisent le fait de créer ?

Une première manière de répondre à cette question est de distinguer, comme l'a fait Margaret Boden dans un article qui fait référence (Boden, 1998), différents types de créativité. Dans cet article, elle distingue en effet la créativité combinatoire, qui consisterait à assembler de manière surprenante différents éléments préexistants, la créativité exploratoire, c'est-à-dire la génération de choses nouvelles par l'exploration d'un espace prédéterminé, et la créativité transformationnelle, qui consiste à modifier cet espace lui-même. Cette troisième forme de créativité est la plus radicale, pourrait-on dire, en ce qu'elle engage des transformations profondes et l'établissement de nouveaux paradigmes. Boden affirme que le troisième type de créativité est le plus difficilement accessible pour une machine. Ces différents types font écho, semble-t-il, à la typologie que nous avons établie : la créativité combinatoire correspond à la corrélation inductive et à la génération aléatoire ; la créativité exploratoire correspond davantage aux réseaux qui donnent une place importante à l'erreur et à l'accident, qui permet de modifier des paramètres dans un cadre établi ; on peut trouver des embryons de créativité transformationnelle dans les dispositifs d'apprentissage par renforcement, qui peuvent produire des phénomènes hors cadre, nouveaux, qui débordent le cadre existant. On pourrait imaginer des dispositifs d'apprentissage par renforcement qui sortent complètement des cadres établis et inventent potentiellement des styles nouveaux, dans la mesure où ils ne sont pas forcément nourris avec des données existantes. La question « une machine peut-elle vraiment créer ? » est donc transformée par Margaret Boden, à partir d'une analyse spécifique des « processus » de la créativité, en une autre question : « Une machine peut-elle utiliser les mêmes processus qu'un individu créatif ? ». À partir de cette reformulation, Boden propose une réponse nuancée, qui varie suivant le type de mécanisme pris en considération.

Aussi stimulante soit-elle, cette manière de poser la question nous semble néanmoins également insuffisante. En effet, on peut tout d'abord souligner que la distinction entre les différentes formes de créativité, qui est en fait une hiérarchisation, puisque la créativité transformationnelle est, semble-t-il, « plus créative » que les autres, peut paraître un peu formelle et extérieure à ce qu'est vraiment un acte de création. Peut-on vraiment affirmer qu'une « grande œuvre transformatrice » est « plus » créative qu'une image poétique particulièrement originale ? Cette partition des types de créativités applique en fait à l'histoire des idées et de l'art une conception kuhnienne (Kuhn, 2008) qui oppose les créations qui s'établissent dans le paradigme normal et les créations qui opéreraient des changements de paradigmes « révolutionnaires », dont il n'est pas sûr qu'elle soit la plus appropriée pour comprendre la notion de création, tant la réflexion de Kuhn donne en fait une place prépondérante aux déterminants sociologiques plutôt qu'au génie individuel. Ensuite, comment déterminer vraiment ce qui relève d'une « simple » combinaison et ce qui relève d'une transformation ? Comment déterminer si une transformation n'est pas au fond une combinaison particulièrement marquante ? Comment déterminer les limites, dont Margaret Boden elle-même explique qu'elles sont floues, entre ce qui relève de l'exploration et ce qui relève de la transformation ? Les œuvres du mouvement Dada, par exemple, peuvent être comprises comme étant des purs jeux de combinaison et pourtant elles ont été profondément transformatrices et ont marqué toutes les avant-gardes du 20^e siècle, donnant naissance au surréalisme par exemple, qu'on aurait également de la peine à classer dans cette nomenclature. On pourrait multiplier les exemples de ce type, mais il ne s'agit pas ici simplement de critiquer la nomenclature de Boden ; toutes les classifications ont leurs limites et il peut paraître trop simple de les critiquer à partir du constat du chevauchement ou de l'indistinction de certaines de leurs catégories. Ce qu'il s'agit plutôt de montrer, c'est que le raisonnement de Boden ne répond pas, au fond, à la question de la machine créative : en effet, même si cette classification avait un sens, ce dont on peut éventuellement douter, elle ne permettrait pas de produire un critère distinctif entre les processus combinatoires, exploratoires et transformationnels qui sont créatifs et ceux qui ne le sont pas. Or la question est précisément d'établir une frontière entre ces deux catégories de processus. Qu'est-ce qu'un processus combinatoire créatif et qu'est-ce qu'un processus combinatoire qui ne l'est pas ? Qu'est-ce qu'une création et qu'est-ce qu'une imitation de la création ? Et c'est la même chose pour les autres processus. Il semble donc qu'il s'agit

d'une erreur de catégorisation : là où nous sommes à la recherche d'un critère de ce que pourrait être quelque chose de créatif, nous rencontrons une catégorisation de processus qui se situe au-delà, ou en deçà, de cette question. Ce qui conduit d'ailleurs Margaret Boden à reproduire, en fait, une analyse relationnelle de la création lorsqu'elle doit répondre à la question de ce qu'est une production créative. Dans son article « Creativity and Artificial Intelligence », elle semble en effet reprendre une définition du nouveau comme étant le surprenant :

There are three main types of creativity, involving different ways of generating the novel ideas. Each of the three results in surprises, but only one (the third) can lead to the « shock » of surprise that greets an apparently impossible idea²⁴ (Boden, 1998).

Plus haut dans son article, elle produit une définition plus complète d'une idée créative. Une idée de ce type serait à la fois nouvelle, surprenante et ayant de la valeur. Il semble au fond que même cette définition ne fasse que reproduire la position « relationnelle » dont nous avons montré les limites : en effet, il semble que 1) « nouvelle » n'est pas défini et que 2) « surprenante » et « ayant de la valeur » sont au fond les deux définitions que nous avons établies plus haut comme étant au fondement de la position relationnelle. Les analyses de Boden, pour stimulantes qu'elles soient, ne semblent donc pas nous permettre de répondre à la question de la machine créative et semblent n'être au fond qu'une autre forme de la position relationnelle.

4.2 Un système déterministe peut-il produire un effet non contenu dans les causes qui le déterminent ?

Il nous semble donc nécessaire d'envisager une dernière manière de poser la question de la machine créative, qui repart d'une analyse du processus lui-même et du rapport entre le déterminisme et la nouveauté. Il s'agit de se demander s'il est possible qu'une machine, c'est-à-dire un système purement déterministe, dont les règles et les arguments sont connus et déterminés à l'avance, puisse produire quelque chose de nouveau, c'est-à-dire un effet qui ne soit pas déjà entièrement contenu dans les causes qui y président.

24. « Il existe trois grands types de créativité, qui impliquent différentes manières de générer des idées nouvelles. Chacun de ces trois types d'idées crée des surprises, mais un seul (le troisième) peut conduire au "choc" de la surprise qui accueille une idée apparemment impossible » (notre traduction).

Cette manière de poser la question de la machine créative pourrait être qualifiée de bergsonienne : pour Bergson, en effet, la création est définie comme l'opposé du déterminisme, puisque, précisément, la nouveauté caractérise le résultat d'un processus dont les effets ne sont pas entièrement contenus dans les causes. Comme il l'explique dans cet extrait de *l'Évolution créatrice*, le calcul ne peut avoir prise sur les phénomènes de création organique, puisque celle-ci, comme toute forme de création véritable, ne peut être réduite à des logiques causales.

De la *création organique*, au contraire des phénomènes évolutifs qui constituent proprement la vie, nous n'entrevoions même pas comment nous pourrions les soumettre à un traitement mathématique. On dira que cette impuissance ne tient qu'à notre ignorance. Mais elle peut aussi bien exprimer que le moment actuel d'un corps vivant ne trouve pas sa raison d'être dans le moment immédiatement antérieur, qu'il faut y joindre tout le passé de l'organisme, son hérédité, enfin l'ensemble d'une très longue histoire (Bergson, 2013, p. 511).

Ce qui rend possible, chez Bergson, la création et le surgissement de la nouveauté, c'est précisément une expérience du temps, sous la forme particulière de ce qu'il appelle la durée. C'est parce qu'un être vivant expérimente le temps comme un moment de changement et d'évolution irréversible qu'il peut y avoir création. Dans un système mécaniste et déterministe, la durée n'existe pas et, au fond, la causalité est une pure identité, il ne peut y avoir de devenir dans un système purement causal. Pour Bergson donc, la machine créative n'a pas de sens, puisque, aussi compliquée soit-elle, elle demeure un système déterministe, dont tous les effets sont contenus dans les causes et qui ne peut produire aucune nouveauté.

Cette position peut toutefois paraître trop univoque en ce qu'elle ne saisit pas la différence qu'il y a entre une machine déterministe de l'âge thermo-industriel et les machines apprenantes que sont aujourd'hui les dispositifs d'intelligence artificielle. Elle ne permet pas de rendre compte du « trouble » que peuvent créer les productions des algorithmes dits créatifs.

Au fond, ce que nous permet de penser Bergson, c'est l'idée que la création procède d'une bifurcation, d'une irruption par rapport au déroulement déterministe de la causalité. Ce qui, à l'origine de la création, est une sortie de la logique purement déterministe, de la logique de la répétition et de pure dérivation. À cet égard, si la notion de durée est évidemment incontournable chez Bergson, elle peut paraître insuffisante, puisque tout ce qui se passe dans la durée n'est pas forcément créatif. Un individu peut

d'ailleurs vivre sur le mode de la machine chez Bergson, vivre comme un automate, dans le régime des habitudes. C'est que la création relève en fait de la rencontre avec un dehors, de l'irruption d'une altérité dans le cours déterministe du même. C'est d'ailleurs ainsi que Deleuze décrit la création et, de manière générale, le travail de la pensée.

Rencontre est le nom d'une relation absolument extérieure, où la pensée entre en rapport avec ce qui ne dépend pas d'elle. L'extériorité des relations est un thème constant chez Deleuze depuis son premier livre. Qu'il s'agisse de penser ou de vivre, l'enjeu est toujours la rencontre, l'événement, donc la relation en tant qu'extérieure à ses termes (Zourabichvili, 1994, p. 23).

La question « une machine peut-elle vraiment créer ? » doit donc recevoir, semble-t-il, une dernière formulation : « Quelles sont les conditions pour créer l'irruption d'un dehors à l'intérieur d'un processus déterministe ? » À cette question, nous ne pouvons apporter ici évidemment une réponse simple et rapide. Mais elle nous paraît plus féconde que les autres manières de poser la question. En effet, elle permet tout d'abord d'ouvrir une perspective nouvelle sur l'histoire de la rationalité. Le dehors du calcul a toujours fait l'objet d'une volonté d'intégration dans le calcul qui lui garderait en même temps son statut de dehors. On peut produire une histoire des formes qu'a prises cette volonté d'intégration du dehors dans le déterminisme. La rationalité calculatoire s'est toujours mise à l'épreuve de son autre pour tenter de le faire sien. Cette histoire commencerait avec les travaux de Leibniz notamment, et sa tentative d'établir un *ars inveniendi*²⁵, c'est-à-dire un calcul permettant l'invention et qui mènerait jusqu'à la deuxième cybernétique, autour de la notion de rétroaction (*feedback*) et d'autopoïèse. Il existe donc des formes particulières, des traductions spécifiques de ce dehors lorsqu'il est intégré dans le calcul. L'intelligence artificielle contemporaine est héritière de ses différentes tentatives, toujours inachevées, que ce soit en laissant une place toujours plus grande au hasard et à l'indéterminé (sous la forme de l'aléatoire) ou en essayant de produire un « inconscient » ou une « imagination » machinique par le libre jeu des calculs²⁶. À cet égard, la notion d'émergence, qui désigne l'apparition de

25. Leibniz a rêvé toute sa vie à l'idée d'un art de l'invention, c'est-à-dire à une sorte de « machine créative » à partir de simples calculs. Cette ambition a pris plusieurs formes, des premiers essais relatifs à la combinatoire jusqu'aux derniers moments d'établissement de la science générale, en passant par les tentatives de caractéristique et d'encyclopédie universelle.

26. On assiste avec l'IA au rêve contemporain de produire une imagination autre que la nôtre, de créer une forme de subjectivité issue des images, une sortie du calcul par le calcul. Voir à cet égard l'intervention de Tristan Garcia lors du séminaire sur l'imagination artificielle (<https://www.youtube.com/watch?v=1jBsxbdFsvg>) et également l'article de Luciana Parisi

quelque chose de fondamentalement nouveau, à un niveau macro, à partir de processus déterministes au niveau micro, a ainsi connu un renouveau théorique et pratique (Cariani, 2008).

Au fond, ce qui compte, c'est moins de savoir si une machine crée ou non une œuvre comme un être humain, mais de savoir quelles ont été les tentatives d'intégrer le dehors dans le calcul tout en lui conservant sa nature de dehors et quelles sont les perspectives à ce sujet. La détermination des manières de le faire ouvre un champ de réflexion sur l'histoire de la rationalité qui pourrait permettre de dépasser les oppositions trop simplistes entre la raison calculatoire d'un côté et ce qui est censé lui échapper de l'autre. Elle peut être à l'origine d'une perspective programmatique, afin de penser une logique algorithmique qui fasse place aux logiques d'enfermement politique, de contrôle et de répétition qui caractérisent aujourd'hui majoritairement l'utilisation de l'intelligence artificielle.

RÉFÉRENCES

- AICAN. *Art of the Future, Now*, <<https://aican.io/>>.
- Bergson, Henri, *L'Évolution créatrice*, Paris, Presses universitaires de France, 2013.
- Boden, Margaret, « Creativity and Artificial Intelligence », *Artificial Intelligence*, vol. 103, n°s 1-2, 1998, p. 347-356.
- Cariani, Peter, « Emergence and Creativity », For the exhibition *Emoção Artificial 4.0 Emergência at Itaú Cultural*, Sao Paolo, Brazil, July 2-September 14, 2008.
- Chatonski, Gregory, « Terre Seconde – Grégory Chatonsky et Emanuele Coccia – Audi Talents » <<https://www.youtube.com/watch?v=90EPfupUsAs>>.
- Danto, Arthur, *La Transfiguration du banal*, Paris, Seuil, 2019.
- Dickie, George, *Art and the Aesthetic*, Ithaca, Cornell University Press, 1974.
- Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny et Marian Mazzone, « CAN : Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms », 2017.
- Fiction de l'artificiel postdigital, <<http://postdigital.ens.fr/>>.
- Jakobson, Roman et Claude Lévi-Strauss, « Les chats » de Charles Baudelaire, dans *L'Homme*, vol. 2, n° 1, 1962, p. 5-21.
- Kuhn, Thomas, *La structure des révolutions scientifiques*, Paris, Flammarion, 2008.
- La lettre du musicien*, « L'intelligence artificielle : la musique du futur », <https://www.lalettredumusicien.fr/s/articles/6398_0_lintelligence-artificielle-la-musique-du-futur>.

- Lausson, Julien, « Pour battre l'IA de Google sur StarCraft 2, vous devez être dans les 0,2% des meilleurs joueurs du monde », *Numerama*, 31 octobre 2019, <<https://www.numerama.com/sciences/565976-vous-ne-pourrez-plus-jamais-battre-lia-de-google-sur-starcraft-2-seuls-les-02-meilleurs-joueurs-resistent.html>>.
- Ourednik, André, *Robopoièses*, Genève, La Baconnière, 2021.
- Parisi, Luciana, « La raison instrumentale, le capitalisme algorithmique et l'incomputable », *Multitudes*, vol. 1, n° 62, 2016, p. 98-109.
- Pasquale, Frank, *The Black Box society : les algorithmes secrets qui contrôlent l'économie et l'information*, Limoges, FYP Éditions, 2015.
- Queneau, Raymond, *Cent mille milliards de poèmes*, Paris, Gallimard, 1961.
- Rees, Marc, « Le chantier de l'exception « text and data mining » au ministère de la Culture », *Nextinpact*, 31 août 2020, <<https://www.nextinpact.com/article/43410/le-chantier-lexception-text-and-data-mining-au-ministere-culture>>.
- Tomas, Anouk, « AlphaFold, l'intelligence artificielle révolutionnaire pour la recherche en protéomique », *Sciences et Avenir*, 2 décembre 2020, <https://www.sciencesetavenir.fr/fondamental/biologie-cellulaire/alphafold-l-intelligence-artificielle-revolutionnaire-pour-la-recherche-en-proteomique_149674>.
- Triclot, Matthieu, *Le moment cybernétique : la constitution de la notion d'information*, Seyssel, Champ Vallon, 2008.
- Turing, Alan, « On Computable Numbers, with an Application to the Entscheidungsproblem », 1936.
- Turing, Alan, « On Computing Machinery and Intelligence », *Mind*, vol. LIX, n° 236, 1950, p. 433-460.
- Weinberg, Justin, « Philosophers on GPT-3 », *Daily Nous*, 30 juillet 2020, <<https://dailynous.com/2020/07/30/philosophers-gpt-3/>>.
- Zourabichvili, François, *Deleuze, une philosophie de l'événement*, Paris, Presses universitaires de France, 1994.

2

FAIRE ŒUVRE AUTONOME ?

CHOMSKY VS. CHOMSKY ET AUTRES ŒUVRES PRODUITES PAR OU AVEC L'IA

Christine Blais¹ et Olivier Blais²

«You could duplicate me, but the question is,
is myself actually myself ? »

CHOMSKY_AI dans *Chomsky vs Chomsky: First Encounter* (2020)

Le monde de l'art contemporain est désormais face à un nouveau phénomène : des œuvres produites à l'aide d'intelligence artificielle (IA) générative ou de réseaux antagonistes génératifs (*generative adversarial networks*, ou GAN³), une technique qui apprend le style d'un corpus d'œuvres afin de s'en inspirer dans la création de nouvelles œuvres uniques. En effet, dans les dernières années, plusieurs œuvres produites *par* une IA ont été exposées et commercialisées sur le marché de l'art spécifiquement sur la base de leur création par une « machine » autonome, alimentant l'intérêt populaire pour ce genre de productions. La possibilité d'un échange fluide

1. Candidate au doctorat en histoire de l'art à l'Université de Montréal.

2. Cofondateur et VP science de la décision chez Moov AI.

3. Le GAN – le terme tant que la technique – a été proposé par Goodfellow (2014) puis formalisé par Radford et ses collaborateurs (2016) pour répondre aux lacunes des modèles génératifs profonds qui reposaient jusqu'alors sur l'approximation de probabilités. L'application de cette technique d'apprentissage-machine (*machine learning*) est particulièrement utile dans la création artistique et linguistique. Le GAN repose sur la mise en compétition de deux réseaux neuronaux, le générateur et le discriminateur, entraînés et semi-supervisés. Le GAN fait maintenant partie de la grande famille de l'IA générative.

avec l'IA et sa capacité à apprendre et à réfléchir sont parmi les promesses les plus intéressantes de l'intelligence artificielle et sont abondamment discutées dans divers domaines des sciences techniques, informatiques et humaines. Une question demeure toutefois moins exploitée dans la sociologie de l'art : si les intelligences artificielles peuvent véritablement générer de manière autonome, deviennent-elles ainsi un acteur légitime du monde de l'art ? L'installation cinématographique en réalité virtuelle *Chomsky vs. Chomsky: First Encounter*⁴ introduite lors du Festival du film de Sundance en 2020, constitue un cas d'étude qui peut offrir une piste de réponse puisqu'elle fut créée avec l'objectif d'alimenter la littérature autour de cette question dichotomique en informant les publics du potentiel et des limites de l'IA, tout en rappelant ce qui est propre à l'intelligence humaine (ONF, 2021). S'inspirant des théories de la grammaire universelle du théoricien Noam Chomsky, l'intelligence conversationnelle au cœur du projet, CHOMSKY_AI, a permis de générer des conversations entre le public et une entité virtuelle à l'image de Chomsky, allant jusqu'à répliquer sa voix et le rythme de sa parole⁵. S'appuyant sur la position du penseur, l'IA apparaît être condamnée à être guidée par des paramètres stricts commandés par l'humain et, ainsi, incapable de l'autonomie technique requise pour une création légitime. Cet essai détaillera la mécanique de création de *Chomsky vs Chomsky: First Encounter* (2020) du point de vue des sciences des données, puis la seconde partie théorique permettra de nuancer certaines préconceptions, notamment en matière d'autonomie, et de souligner les enjeux éthiques et économiques soulevés par les œuvres produites « par » des intelligences artificielles créatives qui circulent dans les mondes de l'art.

4. L'œuvre présentée à Sundance en janvier 2020 était un prototype en réalité virtuelle à utilisateur simple (VR single user). Le prototype se nommait *Chomsky vs Chomsky: First Encounter* et il s'agit de la version de l'œuvre sur laquelle s'appuie ce texte. Une seconde itération fut lancée à Berlin en novembre 2022 puis à Montréal en septembre 2023. Il s'agit d'une installation multi-utilisateur en réalité virtuelle intitulée *CHOMSKY vs CHOMSKY: une curieuse conversation sur l'IA*. Pour plus d'informations, voir le dossier de presse de l'ONF en ligne : ONF, « Chomsky vs Chomsky. Une curieuse conversation sur l'intelligence artificielle », <https://espacemedia.onf.ca/epk/chomsky/?_gl=1%2As0h63c%2A_ga%2AMTM00Tk1MDE4OC4xNjY3MjQwNTg1%2A_ga_EP6WV87GNV%2AMTY3MTYzMTCyNC45LjEuMTY3MTYzMTC0Ny4wLjAuMA>.

5. Il est important de spécifier qu'il ne s'agit pas d'une duplication (communément nommée un *deep fake*). L'intention artistique derrière ce projet est de créer une rencontre avec « une entité Intelligence artificielle » inspirée du célèbre penseur et linguiste.

1. ÉTUDE DE CAS : CHOMSKY VS CHOMSKY : FIRST ENCOUNTER (2020)

Après avoir enfilé des lunettes de réalité virtuelle, les visiteurs du festival du film de Sundance en 2020 avaient la possibilité de rencontrer Noam Chomsky, célèbre penseur et linguiste, pour des discussions profondes et philosophiques ou, plutôt, une entité numérique capable de répondre à toutes les questions qui lui étaient posées à la manière du théoricien. Cette « entité Intelligence artificielle » est autonome dans ses réflexions et capable d'une forme de pensée humaine dans la construction de phrases qui répliquent parfaitement le ton de voix, le style et le rythme de la source originale. Coproduite par l'Office national du film du Canada (ONF), EyeSteelFilm et le collectif d'art médiatique allemand Schnelle Bunte Bilder, cette rencontre consiste en une expérience à tous les niveaux : expérientielle et immersive pour le spectateur désormais participant grâce au dispositif de visionnement, mais également expérience scientifique menée par la sociologue des technologies des médias et réalisatrice Sandra Rodriguez, en collaboration avec plusieurs experts dont le MIT Media Lab du Massachusetts Institute of Technology (MIT), Schnellebuntebilder, studio créatif allemand qui a assuré la direction artistique, le collectif d'artistes visuels berlinois KlingKlangKlong chargé de la création sonore immersive, et la firme de consultants montréalaise en intelligence artificielle Moov AI.

Pourtant, le linguiste est un détracteur notoire de l'IA, de son potentiel d'intelligence et d'autonomie (Biggs 1994 ; Lyre 2020 ; Sampson 2018 ; Kuo 2020). Connu des cercles universitaires depuis des décennies comme le père de la linguistique moderne, et d'un public plus élargi pour ses nombreuses sorties médiatiques à propos de la politique et de l'actualité, Noam Chomsky s'est, dès les années 1950, formellement opposé à la théorie dominante soutenue par l'école des behavioristes voulant que l'humain apprenne le langage au moyen du renforcement. Une de ses théories phares, la grammaire universelle, s'appuie sur le fait que non seulement le renforcement n'est pas la seule clé pour l'apprentissage du langage, mais plutôt qu'une structure universelle de génération du langage doit nécessairement être préexistante. Ses théories fondatrices en sciences cognitives et en linguistique le positionnent forcément aux antipodes du champ de l'intelligence artificielle qui fonctionne sur la base du renforcement. Chomsky s'est d'ailleurs montré publiquement sceptique des techniques d'apprentissage automatique de l'IA en 2012 à l'occasion d'un forum organisé pour le 150^e anniversaire du MIT, critiquant la réminiscence du behaviorisme dans

le domaine de l'IA qui fait généralement usage de techniques statistiques et de masses de données dans une logique prédictive, ce qui serait contre les principes généraux de la cognition.

C'est justement là que réside l'intérêt du projet multifacettes *Chomsky vs Chomsky: First Encounter* : l'IA est ici considérée comme un outil créatif qui démontre le potentiel expressif unique de l'humain. Cette expérience immersive et interactive repousse les frontières de la technologie et du cinéma en jouant avec les codes des fictions dystopiques qui, historiquement, dépeignent la machine comme un être artificiel surpassant l'humain⁶. Nous adhérons plutôt à la position de Chomsky qui compare l'intelligence générale de l'IA à une grue au sommet d'un gratte-ciel : certes impressionnante mais consistant finalement en un outil avec des ancrages fixes (Chomsky cité dans Kuo, 2020). Si l'objectif de cet essai n'est pas de se positionner dans ces débats, l'étude de l'œuvre *Chomsky vs Chomsky: First Encounter*, de son processus créatif et du paramétrage de l'intelligence conversationnelle au cœur du projet permettra d'illustrer le décalage entre l'idéologie hypermédiatisée de l'IA et son infrastructure.

1.1 Construction de l'intelligence conversationnelle CHOMSKY_AI

L'apprentissage automatisé est une des tendances du champ de l'intelligence artificielle comprise dans l'apprentissage profond (*deep learning*), rendu possible par les possibilités techniques des réseaux neuraux artificiels (*artificial neural network*, ou AAN). Les promesses de ces techniques sont de développer de « riches » modèles hiérarchiques représentant la distribution de probabilités sur les données rencontrées par les usages de l'intelligence artificielle, telles que les images naturelles, les motifs du langage ou les symboles des diverses langues » (Goodfellow *et al.*, 2014, p. 1). L'analyse du processus de construction et de fonctionnement de l'intelligence conversationnelle au cœur de l'installation utilisant la réalité virtuelle et un modèle d'IA générative, *Chomsky vs Chomsky: First Encounter*,

6. Cette ambiguïté entre le caractère « naturel » et « artificiel » des machines modernes aux caractéristiques s'inspirant de celles de l'humain fut abordé par des théoriciens de diverses traditions demeure donc, avec les « artistes IA », un enjeu entourant la légitimité de la technologie comme artiste. Sans valider la thèse cybernétique de l'intelligence générale de l'IA, comparable à celle des humains, des travaux empiriques et théoriques, de la sociologie au champ des sciences et technologies (STS) remettent en effet en cause cette opposition entre « naturel » et « artificiel », « réel » et « construit », « humain » et « machine » – pensons entre autres à Bruno Latour, Michel Callon, Steve Woolgar, John Law, Michael Lynch et de nombreux autres.

illustre le processus créatif derrière l'œuvre. Dans l'explication suivante, le rôle de l'expert est habituellement pris par un expert en mégadonnées ou un artiste qui a des aptitudes en science des données.

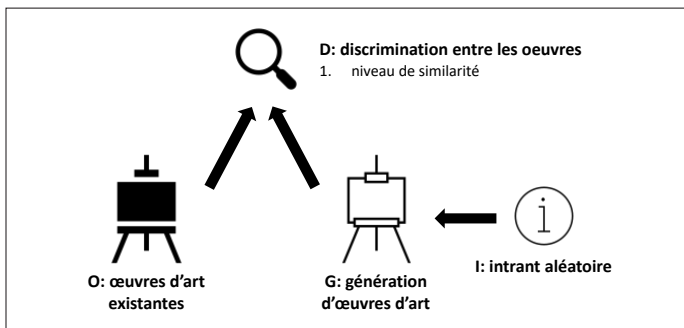
1.1.1 Trouver la stratégie de création

Premièrement, l'expert technique doit trouver la stratégie la plus appropriée pour atteindre les objectifs anticipés par l'artiste et qui peuvent être différents d'un projet à l'autre. Typiquement, ces objectifs tournent autour de l'automatisation d'une tâche habituellement réservée aux humains, ou de la création de contenu de façon non supervisée.

Dans le cas de *Chomsky vs Chomsky : First Encounter*, l'objectif était d'entretenir une conversation de façon complètement autonome avec les spectateurs, à la manière de Noam Chomsky à partir des traces numériques du théoricien. Pour ce faire, un modèle de génération du langage a été conçu par Moov AI, consistant en une adaptation d'un modèle nommé « GPT-2 ». Ce modèle consiste à prédire le mot subséquent d'une séquence de mots en boucle afin de générer une phrase ou une série de phrases (Lin *et al.*, 2020).

Dans le cas de projets plus communément discutés au sein de la communauté artistique, on pourrait parler de la génération autonome d'œuvres d'art, comme ce fut le cas avec le *Portrait d'Edmond de Belamy* (2018), ou encore d'éléments ou de motifs contenus dans une œuvre ou une installation, comme le font certains artistes tels Memo Akten ou Gregory Chatonsky. Pour répondre à ces objectifs, les experts exploitent habituellement des modèles se trouvant dans la famille des GAN (Jabbar *et al.*, 2020).

FIGURE 2.1
Fonctionnement des GAN



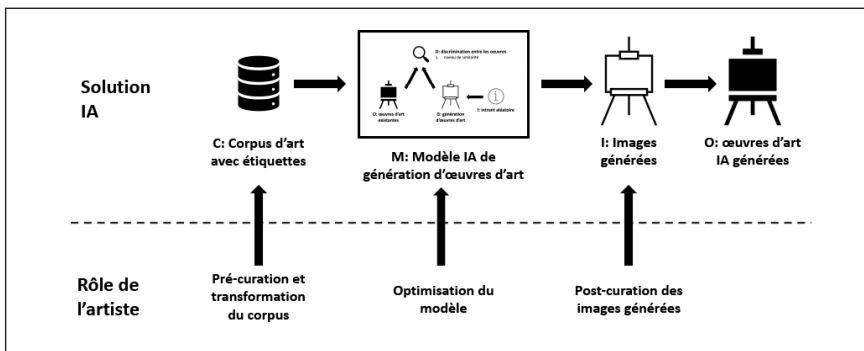
Comme le montre l'image en figure 2.1, un GAN fonctionne en deux étapes :

1. un générateur d'image (G) engendrera des images aléatoires générées de toute pièce, c'est-à-dire que quelques pixels seront remplis aléatoirement ;
2. les images générées sont alors comparées à des images réelles dans le but de les discriminer (D), c'est-à-dire les valider pour ainsi inciter le générateur à créer des images qui sont semblables à un corpus d'images réelles.

1.1.2 « Pré-curation » des données

Lorsque la stratégie appropriée a été trouvée, l'expert doit choisir le bon corpus de données pour être en mesure d'atteindre les objectifs. La suite se déploie en trois phases de traitement et d'organisation des données, comme nous l'illustrons dans la figure 2.2 : « pré-curation », optimisation du modèle et « post-curation ».

FIGURE 2.2
Processus de création avec l'IA et rôle de l'artiste



Dans le cas du projet, l'expert a eu la responsabilité de trouver des milliers de questions manuscrites d'entrevues posées à Chomsky, ainsi que les réponses associées. Il était essentiel que les réponses soient associées aux questions pour que l'IA soit en mesure d'interpréter une question et d'y répondre de façon appropriée. De plus, il était impératif que les réponses proviennent de Noam Chomsky dans le but d'imiter son style linguistique.

Plus largement, dans le cas de la génération d'œuvres d'art, la phase de « pré-curation » (ou la collecte de données) consiste en la création d'un

corpus assez large d'œuvres d'art ou d'images. Dans certains cas, l'expert décidera de restreindre le corpus à des œuvres ou des images similaires qui ont quelque chose en commun (par exemple, un style ou une thématique similaire). Dans d'autres cas, le corpus peut être plus diversifié, mais il serait habituellement accompagné d'étiquettes qui décrivent le style ou d'autres éléments des images du corpus. La décision de restreindre ou d'étiqueter les images contenues dans le corpus par l'expert a pour but de faciliter l'apprentissage des styles ou de certaines thématiques par le modèle (Chen *et al.*, 2020).

1.1.3 Transformer les données

Lorsque les données ont été collectées, il est temps pour l'expert de les transformer afin qu'elles puissent être exploitées par le modèle choisi. Dans l'exemple de *Chomsky vs Chomsky : First Encounter*, les questions et réponses utilisées étaient en anglais et limitées aux caractères alphanumériques et à la ponctuation de base. Finalement, le format des questions et des réponses a été transformé pour satisfaire les critères du modèle GPT-2. Le format de chaque question et réponse ressemblait à ceci :

<Q> question 1 <A> réponse 1 <fin> <Q> question 2 <A> réponse 2 <fin> ...

Dans des cas de génération autonome d'œuvres d'art, l'expert devra convertir chaque pixel des images du corpus en valeur *rouge vert bleu* (RGB), créant une matrice de valeurs numériques.

1.1.4 Entraîner le modèle

Lorsque les données sont disponibles et dans le bon format, l'expert doit maintenant entraîner les modèles IA. La phase d'entraînement consiste à optimiser un modèle pour qu'il s'adapte au corpus. Un modèle peut apprendre grâce à des paramètres qui conservent de l'information à propos de tendances et d'associations qui permettent de générer les prédictions ; il s'agit d'une forme d'apprentissage par renforcement.

Dans la création de CHOMSKY_AI, l'expert a adapté le modèle GPT-2 (*GPT2-medium* pour être exact), déjà entraîné sur un corpus de textes très variés (Radford *et al.*, 2019) avec le corpus précédemment défini. La phase d'entraînement a permis d'optimiser les 355 millions de paramètres du modèle pour qu'ils soient en mesure de prédire le plus précisément possible le prochain mot utilisé considérant les mots précédents. Bien qu'il soit impossible de définir explicitement les paramètres retenus par

l'algorithme puisqu'il s'agit essentiellement de centaines de millions de valeurs numériques pondérées à l'aide d'une équation mathématique, il est toutefois possible d'affirmer que GPT-2 exploite surtout les associations entre différents mots, l'attention portée sur certains de ceux-ci dans certains contextes et la génération de structures de phrases. Pour la génération d'œuvres d'art, la même phase d'entraînement est requise dans le but d'optimiser la génération d'œuvres similaires au corpus retenu. Le processus d'entraînement consiste également à générer un grand volume de paramètres à régler.

1.1.5 La « *post-curation* »

Durant le processus d'entraînement, l'expert testera une multitude de configurations de modèles, communément appelé des « hyperparamètres ». C'est lorsque l'expert aura analysé la qualité de l'extrait qu'il sera assez à l'aise pour sélectionner le modèle final. En d'autres mots, l'expert a accès à des métriques pour quantifier la performance de chaque configuration de modèle. Par exemple, il est possible d'avoir des statistiques qui décrivent à quel point un modèle de génération du langage ou de génération d'œuvres d'art est capable de reproduire un élément provenant du corpus. Cependant, la véritable validation se fait lorsque l'expert tente de générer de nouvelles phrases ou œuvres, puisqu'un modèle peut être efficace pour reproduire des éléments du corpus sans avoir la stabilité pour produire de nouveaux éléments de bonne qualité (Pasini et Baralis, 2019). Cette stabilité acquise par le modèle se traduit par une autonomie potentiellement accrue.

1.2 Paramétrages ou limites de l'influence humaine

Un des défis que l'expert a eu à surmonter dans l'élaboration de CHOMSKY_AI était de créer un modèle qui était capable de générer des réponses relativement longues, intéressantes et cohérentes à propos d'une multitude de thématiques, sachant que le chercheur nonagénaire est particulièrement prolifique en matière d'écrits, de conférences et d'entrevues. Lors d'un projet de génération d'œuvres d'art, l'expert devra s'assurer que l'œuvre est intéressante sans toutefois copier une œuvre existante. De plus, l'expert ou l'artiste aura potentiellement des critères au-delà de l'esthétisme de l'œuvre tels que le respect d'un style ou d'une thématique qui

s'inscrit dans un concept artistique. Voici un exemple de quelques questions posées au double numérique de Noam Chomsky, puis ses réponses :

QUESTIONS	RÉPONSES
What do you think society expects from AI?	There is nothing in the law that says it has to be kind of a god. They are designed by humans who have set out rules for what should happen and how we ought not to handle this problem because if anything goes wrong then things get ugly quickly and people will probably suffer very badly at our hands.
Do animals have language?	There is strong evidence that they can learn to talk and even play music. However we know little about what makes them do it.
Why can't you answer all my questions?	Because these are not the topics we have time for.

Ces réponses sont intéressantes pour plusieurs raisons : premièrement, il ne s'agit pas de copies de réponses existantes de Chomsky, ce qui serait problématique puisque le but de l'expérience est de générer de nouvelles réponses. De plus, les réponses sont logiques et bien écrites, et elles respectent également le style linguistique de Chomsky. Finalement, ces réponses sont pertinentes et captivantes, ce qui est un critère important pour générer une conversation fluide et une expérience enthousiasmante. Par ailleurs, comme nous l'avons vu, le rôle du spécialiste – qu'il s'agisse d'un scientifique ou d'un artiste – est essentiel tout au long du processus de création. C'est d'ailleurs un fait que soutiennent d'autres chercheurs, notamment en regard du processus créatif des œuvres produites à l'aide de l'intelligence artificielle qui implique hautement l'expert ou l'artiste (Elgammal *et al.*, 2017).

En résumé, comme la figure 2.2 le démontre, dans le processus de création d'une œuvre avec l'IA, l'artiste choisit la collection d'images qui nourrira l'algorithme dans la phase de précommissariat (« pré-curation »). Ces images entraîneront l'algorithme (le GAN) qui tentera d'imiter le contenu de sa base de données. Enfin, l'artiste permet l'entraînement de la machine et le développement du modèle par son paramétrage, puis tient le rôle le plus important lors de la sélection finale d'images (« post-curation »). Comme nous le constatons donc en analysant le travail technique requis pour en arriver à générer une œuvre à l'aide de l'intelligence artificielle, tout porte à croire que l'intervention humaine est fondamentale pour en arriver aux résultats escomptés.

2. AUTRES MODÈLES D'IA CRÉATIVES

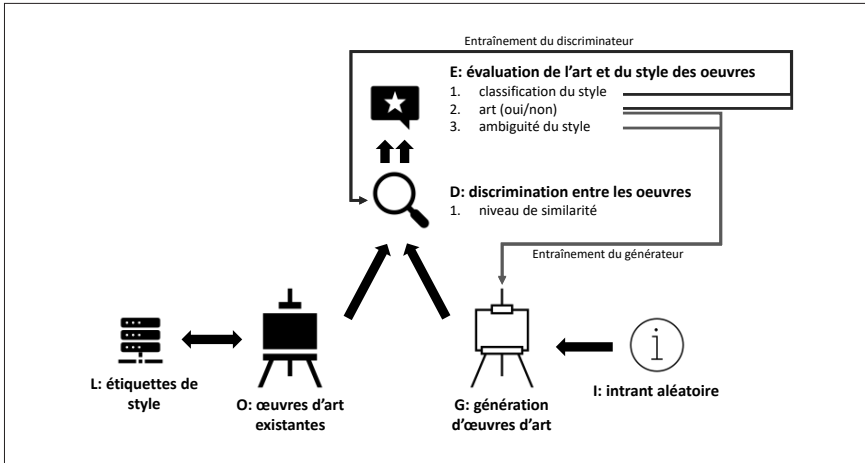
En 2018, Elgammal et ses collaborateurs ont adapté le GAN classique afin de le rendre plus créatif. Ce modèle se nomme le réseau antagoniste créatif (*creative adversarial network*, ou CAN) et permet d'atteindre des objectifs similaires au GAN, mais plus efficacement (Elgammal *et al.*, 2018, p. 2). Inspirés par la théorie de Martindale à propos de la nouvelle création en arts visuels, c'est-à-dire qu'elle doit détenir un haut potentiel de stimulation afin d'être en mesure de repousser les limites, sans trop dévier des conventions esthétiques, Elgammal et son équipe ont proposé un réseau structuré afin de générer de l'art qui ne suit pas les mouvements artistiques ou styles établis, mais qui tente plutôt de générer de l'art qui confond le plus possible le spectateur humain (Elgammal *et al.*, 2018, p. 5). Le CAN proposé se distingue également des GAN traditionnels en étant moins émulateur et davantage créatif. Cela est possible par le générateur qui est conçu pour recevoir deux signaux des discriminateurs (contrairement à un seul signal avec un GAN) qui agissent comme deux forces contradictoires (Elgammal *et al.*, 2018, p. 6). Cette mécanique permet de générer de nouvelles œuvres assez près de l'ensemble de référence, lequel est associé à des styles (baroque, impressionnisme, cubisme, etc.) plutôt qu'à des mots clés spécifiques (portraits, paysages, abstraction, etc.), mais sans trop s'en écarter, auquel cas cela activerait un système d'aversion. Enfin, les œuvres générées devraient augmenter l'ambiguïté stylistique (Elgammal *et al.*, 2018, p. 6). AICAN serait donc plus qu'un simple outil technologique puisqu'il choisit « ses propres thèmes, expérimente et apprend sans intervention humaine » (Godin-Filion, 2020).

La différence entre un GAN et un CAN se situe donc à deux niveaux :

1. Plus de données d'entraînement sont requises puisque le modèle n'a pas besoin seulement du corpus d'œuvres d'art que le GAN classique nécessite, mais aussi des étiquettes associées aux types d'œuvres ainsi qu'aux styles des œuvres. Il est bon de savoir que certains modèles de GAN tels que le « *Condition GAN* » (Jabbar *et al.*, 2020) peuvent aussi utiliser des étiquettes similaires. Cela permet essentiellement de guider le processus créatif en donnant des instructions lors de la création des œuvres ;
2. La couche discriminante décrite à l'étape 2 du GAN comparera les œuvres réelles et générées à l'aide de critères plus précis que ceux du GAN, notamment en distinguant si l'œuvre générée est censée être de l'art, et en s'assurant du respect des styles artistiques.

Selon les recherches effectuées sur le CAN, l'effet de cette discrimination permet au modèle de générer des œuvres artistiques tout en respectant les styles artistiques.

FIGURE 2.3
Processus d'entraînement de l'IA



En somme, le CAN proposé par Elgammal et son équipe est novateur à certains niveaux du point de vue technique, car il propose un système qui apprend à propos de l'histoire de l'art à l'intérieur même de son processus créatif, bien qu'il n'ait aucune compréhension sémantique de l'art et des concepts de styles. Le système aurait, en théorie, la capacité de continuellement apprendre à créer de nouvelles œuvres dans la lignée de l'histoire de l'art, mais sans copier un style ou un thème particulier, et d'adapter sa génération sur la base de ce qu'il a appris. Si AICAN est aujourd'hui présentée comme une « IA artiste », on peut se demander dans quelle mesure elle est entièrement autonome. Elle peut être créative, mais les paramètres ont été pensés, prévus et programmés par des humains. C'est d'ailleurs une observation que fait Bernier (2019) en soulignant que l'algorithme développé par Elgammal et son équipe est basé sur une définition de créativité qui l'assimile totalement à celle d'originalité, un amalgame nettement plus complexe qu'il le décrit et qui fait par ailleurs l'objet de remises en question au sein même de la discipline de l'histoire de l'art (p. 16).

Il serait également possible de retenir la proposition de l'artiste, ingénieur et expert de mégadonnées Memo Akten qui, non seulement met en

relief les failles de l'IA dans sa pratique artistique, mais a proposé une nouvelle méthode pour l'utilisation d'un GAN, conçu pour sa série *Learning to See* (2017). Cette série, qui prenait la forme d'une installation interactive, démontrait comment un réseau neuronal « voit le monde » et tente de donner un sens à ce qu'il voit, c'est-à-dire à travers le filtre de ce qu'il connaît déjà, « comme nous » précise l'artiste (Akten, 2018). Présentée dans de nombreux établissements muséaux, l'installation se décline ainsi : des réseaux neuraux analysent une captation vidéo en direct, pointant à une table couverte d'objets usuels. Le public est invité à manipuler ces objets et leur configuration est projetée sur un écran. Les réseaux neuraux analysent en temps réel la configuration du public et affichent la réinterprétation (les prévisions du GAN) sur un écran juxtaposé. Cette installation explore les biais dans les réseaux neuraux artificiels et la manière dont les humains et les « machines » construisent du sens. Elle propose également des mécanismes pour la manipulation de représentations spécifiquement entraînées pour le monde réel (Akten *et al.*, 2020, p. 1), en temps réel. Pour revenir sur l'explication du rôle de l'expert tenu précédemment, il est nécessaire de souligner la particularité de l'interaction en direct dans le projet d'Akten qui implique une opérationnalisation du modèle IA sur des infrastructures de technologie de l'information (TI). Plus précisément, opérationnaliser un modèle consiste à rendre un modèle disponible en ligne ou sur un serveur afin qu'il soit capable de générer des prédictions lorsque les spectateurs interagissent avec la solution. Pour le projet *Chomsky vs. Chomsky*, la tâche d'opérationnalisation a pris près de 30 % du temps de création du modèle IA, puisqu'il a été nécessaire d'établir un système capable de capter des questions vocales, de les convertir en texte, de générer une réponse à l'aide du modèle IA décrit précédemment puis de retourner une réponse en imitant la voix de Noam Chomsky.

Dans *Learning to See* (2017), la technique utilisée pour mettre en œuvre ce mécanisme a été mise au point par Akten et deux collègues chercheurs. Il s'agit d'un système d'entraînement et d'un logiciel personnalisé facilitant une interaction ludique et en continu pour un contrôle humain significatif, tout en offrant un éventail de possibilités créatives non disponibles dans aucun autre système, selon Akten et ses collaborateurs. Cet instrument est décrit comme ceci : « Un réseau neural artificiel basé sur les images, entraîné sur un nombre de différents ensembles de données personnalisés, effectuant des prévisions à partir des intrants de la captation vidéo en direct » (Akten *et al.*, 2020, p. 1). Conceptuellement, la motivation derrière cette œuvre est d'utiliser ces réseaux neuraux convolutifs de manière très

libre et inspirée par notre propre cortex visuel, en tant que réflexion sur notre propre affirmation des biais cognitifs (Akten *et al.*, 2020, p. 2).

Le type de pratique que nous venons de décrire est au cœur de l'émergence de nouvelles politiques dans ce que Cubitt et Thomas (2015) décrivent comme la « modestie de l'esthétique relationnelle : le « remake » des politiques pour incorporer nos environnements techniques et naturels » (p. 37). Dans *Learning to See*, où les utilisateurs manipulent directement avec les mains, le corps ou les objets en face d'une caméra, l'interaction entre l'utilisateur, le contenu visuel et le résultat (*output*) est si immédiat que l'utilisateur est à même d'expérimenter, d'improviser et d'améliorer le contenu visuel. *Chomsky vs Chomsky: First Encounter* et *Learning to See* comportent une similarité importante : le regardeur devient participant, voire co-créateur de l'œuvre, puisqu'il est en interaction avec l'IA et interagit en direct sur le résultat obtenu. Comme Akten et ses collaborateurs (2020) le soutiennent, peu de recherches offrent de réponses sur le potentiel de l'apprentissage profond en tant qu'instrument ou outil collaboratif humain-machine en temps réel⁷ (p. 3). Et c'est là que réside aussi l'une des avancées proposées dans *Chomsky vs Chomsky: First Encounter*.

3. LA COLLABORATION AVEC L'IA

Comme nous l'avons vu, la manipulation humaine est nécessaire à tous les stades de la manipulation des données. À cet égard, la plupart des artistes nomment l'IA comme un *collaborateur* de leur œuvre : si l'accès aux différentes techniques et systèmes de l'IA en accès ouvert sur Internet permet une démocratisation de cet outil, d'autres privilégieront des collaborations avec des spécialistes de ces technologies.

Il est toutefois nécessaire de soulever que certains considèrent que l'IA n'est pas qu'un outil⁸, mais plutôt une cocréatrice avec l'artiste⁹. Si cette étiquette « d'artiste » sur un système d'IA est débattable, il n'en demeure

7. Dès 1999, Peter Weibel suggérait qu'avec la possibilité de contenu généré par le consommateur il était possible d'observer une nouvelle forme d'art démocratique dans laquelle tout le monde pouvait participer.

8. Comme le soutient d'ailleurs Hertzmann (2018) : « Tous les algorithmes en art, incluant les méthodes d'apprentissage machine, sont des outils pour les artistes ; ils ne sont pas des artistes eux-mêmes » (p. 16).

9. Tentant au contraire de démontrer la capacité totale de création d'AICAN, Elgammal affirme plutôt : « Si vous avez des artistes travaillant dans votre atelier et cocréant avec vous, les décririez-vous comme des outils ? » (Elgammal cité dans Godin-Fillion, 2020).

pas moins que l'outil prend désormais une place plus importante dans le processus créatif, car il est, avec les techniques d'apprentissage automatique, capable de surprendre l'artiste en créant quelque chose auquel il ne s'attendait pas¹⁰. Il est toutefois important de rappeler que la perspective des chercheurs à l'origine d'AICAN présente la créativité de l'IA comme un résultat objectif, alors que la démarche de l'artiste dans la production de son œuvre est nécessairement le produit de la créativité humaine, « dans toute sa subjectivité » (Bernier, 2019, p. 19).

Justement, Chomsky établissait dans son ouvrage de 1966, réédité en 2009, que l'idée généralisée du pouvoir cognitif (« esprit »), ne prenant son sens qu'en présence d'une certaine forme de création, est une notion développée dès le 16^e siècle et discutée par Descartes notamment en regard de la « machine » :

[T]he idea that the « cognitive power » is properly called « mind » only when it is in some sense creative has earlier origins. One source that might well have been familiar to Descartes is Juan Huarte's Examen de Ingenios (1575), which was widely translated and circulated (I quote from the English translation by Bellamy, 1698). Huarte understands the word Ingenio to have the root meaning « engender », « generate » – he relates it to gigno, genero, ingenero. [...]. Thus « one may discover two generative Powers in Man, one common with the Beasts and Plants, and the other Participating of Spiritual Substances, God and the Angels » [...]. « Wit [Ingenio] is a generative power... the Understanding is Generative Faculty » [...]. As distinct from divine « Genius », the human « rational soul » and « spiritual substances » do not have sufficient Force and Power in their Generation to give real being to what they Ingender » but only « to produce an accident in the Memory, » « an Idea and Image of what we know and understand » that must be given concrete existence by work and art (Chomsky, 2009, p. 121).

Ce raisonnement humaniste voulant que certains apprentissages, rationnels ou spirituels, s'inscrivent dans l'esprit « pour produire un accident dans la mémoire, une idée et une image de ce que l'on connaît et comprend » (notre traduction), fait pourtant étrangement écho au mode de fonctionnement et à l'utilisation de l'intelligence artificielle et conversationnelle en arts visuels contemporains. À la différence, comme nous l'avons vu, que le pouvoir génératif d'un langage (visuel ou textuel) des GAN et des CAN est alimenté par l'humain, qui nourrit la « mémoire » de cette technologie à l'aide d'une banque de référence (textes, images, etc.)

10. Elgammal (2018) lui-même tomba dans ce piège en lançant, en conclusion d'une conférence TED : « Am I an artist since I programmed the machine ? ».

canonisée et paramétrée. Au fil de son apprentissage, la « machine », comme l'humain, fera certaines erreurs – des accidents que l'on nomme des *glitches* dans les produits numériques – à partir desquelles certains diront que la machine apprendra afin de toujours s'améliorer.

De l'autre côté, la collaboration avec des entreprises spécialisées hors du réseau culturel peut être requise pour mettre en œuvre certaines créations qui requièrent un haut niveau de compétences techniques, comme ce fut le cas avec *Chomsky vs Chomsky: First Encounter* et CHOMSKY_AI. Ce type de collaboration entreprises-artistes n'est pas un nouveau phénomène, mais elle s'élabore également avec les technologies de manière cocreative, ce qui est au cœur des questions sur les œuvres créées avec l'IA. Toutefois, il est nécessaire de souligner que, comme l'indique le sociologue de l'art Jean-Paul Fourmentraux (2007), « la cocreation n'entraîne pas automatiquement la covalidation » : si la pratique de l'art peut relever d'une activité collective, les contributions à la production laissent historiquement la place à la « figure mythique de l'auteur », qui est garant de la légitimité de l'œuvre. En outre, même s'il est appuyé par un « personnel de renfort », le schéma traditionnel des mondes de l'art défini par le sociologue Howard Becker en 1989 fait de l'artiste le créateur exclusif de ses œuvres. Il s'agit là d'arguments en défaveur d'une autonomie de la machine, défendant plutôt la coexistence des systèmes et de l'humain dans le champ de l'intelligence artificielle (Lovelock, 2019 ; Akten, 2019). On observe donc la distinction sémantique entre ce qui incombe au champ de recherche et les caractéristiques de l'IA : les propriétés qui définissent l'intelligence que l'on paramètre dans des systèmes (apprentissage, recherche, orientation, etc.) ne rendraient pas ces systèmes intelligents en soi.

4. ENJEUX ÉTHIQUES ET ÉCONOMIQUES DANS LE MARCHÉ DE L'ART : ALLER AU-DELÀ DE LA CAPACITÉ CRÉATRICE DE L'IA

Comme nous l'avons vu précédemment, la créativité n'est pas le seul critère d'authenticité d'une œuvre. En effet, dans le monde de l'art contemporain, la légitimation d'une œuvre s'effectue notamment à travers l'intentionnalité, la « touche » de l'artiste, son *aura*, ou sa réception et sa circulation dans le circuit institutionnel et marchand. La vente record du *Portrait d'Edmond de Belamy* chez Christie's en 2018, décrite par la prestigieuse maison de vente aux enchères comme produit par une IA générative, a marqué un tournant important en cela qu'une œuvre créée par une

intelligence artificielle était désormais acceptée des réseaux traditionnels du marché de l'art. En outre, elle fut adjugée pour l'importante somme de 432 500 dollars américains. Puis les œuvres produites par AICAN ont été présentées au grand public aux importantes foires d'art contemporain Scope de New York et Art+ à Francfort en 2018, ainsi que lors de l'exposition *Faceless Portraits Transcending Time*, du 12 février au 5 mars 2019 à la galerie HG Contemporary de New York. Dans l'exposition en galerie, elles étaient présentées et commercialisées comme le produit d'un « artiste intelligence artificielle et un partenaire créatif » (Elgammal *et al.*, 2018; Elgammal et Mazzone 2019).

Le fait que ces œuvres furent exposées dans des contextes institutionnels, soutenus par des acteurs du monde de l'art qui détiennent une position importante dans la construction des valeurs et le fait qu'elles étaient commercialisées comme étant produites *par* une IA sont des facteurs qui soulèvent d'importants enjeux éthiques. En effet, bien que l'algorithme soit un outil relativement nouveau qui intervient désormais dans le fonctionnement des activités collectives de l'art (sa production, sa consommation et sa canonisation), on ne peut affirmer, contrairement à ce qui s'opère dans la sphère marchande, que seule la mise en marché de l'artiste IA valide le statut de ces créations.

Par ailleurs, l'importante financiarisation des œuvres produites *par* l'IA révèle certains enjeux d'ordre économique. En effet, ce phénomène exagère justement l'importante spéculation qui sévit dans le monde de l'art contemporain. Les premiers exemples d'œuvres produites par la machine qui ont intégré le marché de l'art prennent la forme d'œuvres traditionnelles – le plus souvent des œuvres bidimensionnelles et imprimées – « dans un effort de recherche de légitimation » (Doherty, 2019, p. 33). Ce rendu artistique répond à l'enthousiasme du public général et permet aux acteurs du marché de l'art de bénéficier de l'effet de mode pour ce type de production artistique. Cette lecture peut fournir une forme d'éclaircissement sur le phénomène entourant les images soi-disant produites *par* l'IA, dans lesquelles la technologie et la nouveauté sont une plus-value.

Comme nous l'avons vu, une œuvre ou un système dans lequel la capacité d'apprentissage, de reconnaissance vocale ou de création en temps réel a été paramétrée ne devient pas automatiquement intelligent ni créatif. Une supervision par l'expert ou une manipulation des données par l'artiste est requise, bien que sa participation soit moindre avec l'IA générative par la nature de l'apprentissage profond. En somme, l'apprentissage profond

a un grand potentiel à plusieurs égards, incluant celui d'agir comme un partenaire collaboratif dans la production de réalisations créatives, spécifiquement dans une manière continue et en temps réel (Akten, Fiebrink et Grierson, 2020, p. 6). Citons de nouveau Chomsky (2009) qui soutient que la dimension poétique unique du langage dérive justement de son indépendance à stimulation immédiate, une idée soutenue par les écrivains romantiques tels Schlegel et Rousseau. Ainsi, qu'il s'agisse de linguistique ou d'art, les interactions immédiates et spontanées avec l'environnement définissent le potentiel créatif illimité de l'humain (Chomsky, 2009, p. 68). Cela souligne justement les limites de l'algorithme : bien que s'appuyant sur des caractéristiques propres à l'intelligence humaine, cela ne fait pas de l'intelligence artificielle un créateur autonome. Il convient de souligner que, en écho aux propos de Lev Manovich, les conclusions de l'expérience qui entoure la diffusion des œuvres produites grâce à des intelligences artificielles permettent d'affirmer à tout le moins le rôle des sciences des données et de l'apprentissage automatique dans le domaine de l'histoire de l'art en tant que science prédictive qui permet de découvrir les motifs et les tendances fondamentales qui n'appartiennent pas à l'œil humain.

La mécanique de création décrite dans l'étude de cas *Chomsky vs. Chomsky* (2020), ainsi que la discussion autour de la cocréation dans la nouvelle création en arts, démontre que l'IA n'est pas la panacée véhiculée dans la pensée populaire, dans le circuit événementiel ou sur le marché de l'art. Comme il l'a été démontré au cours de cet essai, les conditions de l'autonomie et de la cocréation impliquées de la production, à la diffusion et à la commercialisation de l'œuvre, défendent la position que tout ce qui est bâti autour d'une œuvre composée de systèmes d'intelligence artificielle repose sur des paramètres faits par l'humain. Il demeure que d'importants enjeux d'ordre économique et éthique résident dans la légitimation de ces œuvres par des instances qui ont un pouvoir de validation, sur ce qui est accepté ou non en tant qu'œuvre d'art (les maisons de vente aux enchères, les institutions et les foires et festivals).

Nous avons tenté de démontrer qu'au contraire les œuvres produites avec l'IA permettent autant une réflexion sur la condition humaine dans notre culture numérique, dans laquelle le rôle omniprésent de la technologie permet à l'artiste de se « réapproprier la technologie comme un espace pour l'imagination » (Catricalà, 2020, p. 36). Conceptuellement, parlons plutôt d'un « imaginaire artificiel » : « Entre l'idéologie de l'autonomie et du remplacement, et le redoublement de l'image, s'inscrit la différence

entre l'intelligence et l'imagination artificielles [...]», comme le soutient l'artiste Gregory Chatonsky (2021, p. 37). Et c'est précisément dans cet « imaginaire artificiel » que l'expérience interactive *Chomsky vs Chomsky* et ses itérations futures s'inscrivent.

RÉFÉRENCES

- Akten, Memo, « "Artificial Intelligence" as the name of a research field vs a characteristic of being », *Medium*, 22 juillet 2019, <<https://memoakten.medium.com/artificial-intelligence-as-the-name-of-a-research-field-vs-a-characteristic-of-being-920cb5f3361e>>.
- Akten, Memo, « Learning to See (2017) », 2018, <<http://www.memo.tv/works/learning-to-see/>>.
- Akten, Memo, Rebecca Fiebrink et Mick Grierson, « Learning to See: You are what you see », *Arxiv*, 2020, <<https://arxiv.org/pdf/2003.00902.pdf>>.
- Bernier, Christine, « Muséologie numérique : action participative en réseau et canonisation selon l'intelligence artificielle », *Digital Studies/Le champ numérique*, vol. 9, n° 1, 2019, p. 1-21, <<https://doi.org/10.16995/dscn.346>>.
- Biggs, Colin, « Chomsky and artificial intelligence », dans Carlos Peregrin Otero, *Noam Chomsky: Critical Assessments*, Taylor & Francis, 1994, p. 67-80.
- Catricalà, Valentino, *Art and Technology in the Third Millenial*, Milan et Verone, Electa S.p.A, 2020.
- Chatonsky, Gregory, « Gregory Chatonsky », *Ratsdeville*, 19 juin 2015, <<https://ratsdeville.typepad.com/ratsdeville/2015/06/ratsdevillemontreal-gregory-chatonsky.html>>.
- Chatonsky, Grégory, « La résurrection qui vient », *Écosystème*, vol. 3, 2021, p. 22-54, <<https://ma.x24cdn.com/chambreblanche.qc.ca/media/library/99751/17777/17777.X.pdf>>.
- Chen, Weijie, Shiliang Pu, Di Xie, Shicai Yang, Yilu Guo et LuoJun Lin, « Unsupervised image classification for deep representation learning », Hikvision Research Institute (dir.), arXiv preprint, arXiv:2006.11480v2, School of Electronic and Information Engineering, South China University of Technology, Hangzhou, 2020.
- Chomsky, Noam, *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, Cambridge, Cambridge University Press, 2009 [1966].
- Cohen, Harold, « The further exploits of AARON, painter », *Stanford Humanities Review*, vol. 4, 1995, p. 141-158.
- Colton, Simon, « Creativity versus the perception of creativity in computational systems », *Association for the Advancement of Artificial Intelligence*, 2008, <<https://cdn.aaai.org/Symposia/Spring/2008/SS-08-03/SS08-03-003.pdf>>.
- Cubitt, Sean et Paul Thomas, « The new materialism in media art history », dans Valentino Catricalà, *Media Arts: Towards a New Definition of Arts in the Age of Technology*, Rome, Fondazione Mondo Digitale, 2015, p. 19-38.
- Dennett, Daniel C., « When philosophers encounter artificial intelligence », *Daedalus*, vol. 117, n° 1, 1988, p. 283-295, <<http://www.jstor.org/stable/20025148>>.
- Doherty, Shauna Jean, « Art in the age of artificial intelligence / L'art à l'ère de l'intelligence artificielle », *esse arts + opinions*, vol. 97, 2019, p. 30-41.

- Elgammal, Ahmed, «AI vs art: Can the machine be creative?», *TEDxBratislava*, 2 novembre 2018, <https://www.ted.com/talks/ahmed_elgammal_ai_vs_art_can_the_machine_be_creative/details>.
- Elgammal, Ahmed, «An algorithm for assessing creativity of art», *Medium*, 5 juin 2015, <https://medium.com/@ahmed_elgammal/algorithm-for-assessing-creativity-of-art-a9a8c159082a>.
- Elgammal, Ahmed et Marian Mazzone, «Art, creativity, and the potential of artificial intelligence», *Arts*, vol. 8, n° 1, 2019, <<https://doi.org/10.3390/arts8010026>>.
- Elgammal, Ahmed, Marian Mazzone, Bingchen Liu, Diana Kim et Mohamed Elhoseiny, «The shape of art history in the eyes of the machine», *AAAI*, 2018, <<https://arxiv.org/abs/1801.07729>>.
- Elgammal, Ahmed, Bingchen Liu, Mohamed Elhoseiny et Marian Mazzone, «CAN: Creative adversarial networks generating “art” by learning about styles and deviating from style norms», dans *The Art & AI Laboratory* (dir.), arXiv preprint, arXiv:1706.07068v1, Rutgers University, Piscataway, 2017.
- Elgammal, Ahmed et Babak Saleh, «Quantifying creativity in art networks», Actes de conférence, *International Conference on Computational Creativity (ICCC)*, 6^e édition, juin 2015, <<https://axon.cs.byu.edu/ICCC2015proceedings/2.3Elgammal.pdf>>.
- Fourmentraux, Jean-Paul, «Faire œuvre commune. Dynamiques d’attribution et de valorisation des coproductions en art numérique», *Sociologie du travail*, vol. 49, n° 2, avril-juin 2007, <<https://doi.org/10.4000/sdt.21041>>.
- Godin-Filion, Vincent, «AICAN: l’artiste créé, entrevue avec Dr Ahmed Elgammal», *IA CIAC MTL*, n° 3, 15 avril 2020, <<http://ciac.ca/ia-ciac-mtl-03-02/>>.
- Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville et Yoshua Bengio, «Generative adversarial networks», *Advances in Neural Information Processing Systems 27*, 2014, <<https://arxiv.org/abs/1406.2661>>.
- Haraway, Donna, «A Cyborg Manifesto: Science, technology, and socialist feminism in the late twentieth century», dans Donna Haraway, *Simians, Cyborgs and Women: The Reinvention of Nature*, New York, Routledge, 1991, p. 149-181.
- Heidenreich, Hunter, «What is a generative adversarial network?», billet de blogue, 18 août 2018, <http://hunterheidenreich.com/blog/what-is-a-gan/?fbclid=IwAR1ZCTBVLjks7BDe_cdurV_Tv7AKWtVZRFEve_kXSfArTv3wGioWuArtCzo>.
- Hertzmann, Aaron, «Can computers create art?», *Arts*, vol. 7, n° 2, mai 2018, <<https://arxiv.org/abs/1801.04486>>.
- Jabbar, Abdul, Xi Li et Bourahla Omar, «A survey on generative adversarial networks: Variants, applications, and training», dans *Hikvision Research Institute* (dir.), arXiv preprint arXiv:2006.05132, College of Computer Science, Zhejiang University, Hangzhou, 2020.
- Kuo, Andrew, «Noam Chomsky on the future of deep learning», *Towards Data Science*, 5 décembre 2020, <<https://towardsdatascience.com/noam-chomsky-on-the-future-of-deep-learning-2beb37815a3e>>.
- Latour Bruno, «La fin des moyens», *Réseaux*, vol. 18, n° 100, 2000, p. 39-58.
- Lin, Zhaojiang, Andrea Madotto et Pascale Fung, «Exploring versatile generative language model via parameter-efficient transfer learning», dans *Center for Artificial Intelligence Research (CAiRE)* (dir.), arXiv preprint, arXiv:2004.03829v2, The Hong Kong University of Science and Technology, Hong Kong, 2020.

- Lovelock, James, *Novacene : The Coming Age of Hyperintelligence*, Cambridge, Massachusetts, The MIT Press, 2019.
- Lyre, Holger, « The state space of artificial intelligence », *Minds and Machines*, vol. 30, 2020, p. 325-247.
- Mitchell, W. J. T., « La plus-value des images », *Études littéraires*, vol. 33, n° 1, 2001, p. 201-225, <<https://doi.org/10.7202/501286ar>>.
- Nicholson, Chris, « A beginner's guide to generative adversarial networks (GANs) », *Pathmind*, 2020, <https://wiki.pathmind.com/generative-adversarial-network-gan?fbclid=IwAR2yTQaSx8m_hQIfEKKwp58JqJ0HmGPlz-XqTydiMrGwHJPDIGo-4j9yk30>.
- ONF, *Chomsky vs. Chomsky : First Encounter*, Office national du film, 2021, <https://www.nfb.ca/interactive/chomsky_vs_chomsky/>.
- ONF, « Chomsky vs Chomsky. Une curieuse conversation sur l'intelligence artificielle », dossier de presse, Office national du film, 2022, <https://espacemedia.onf.ca/epk/chomsky/?_gl=1%2As0h63c%2A_ga%2AMTM00Tk1MDE4OC4xNjY3MjQwNTg1%2A_ga_EP6WV87GNV%2AMTY3MTYzMTcyNC45LjEuMTY3MTYzMTc0Ny4wLjAuMA>.
- Pasini, Andrea et Elena Baralis, « Detecting anomalies in image classification by means of semantic relationships », *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, Italie, 2019, p. 231-238.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei et Ilya Sutskever, « Language models are unsupervised multitask learners », *OpenAI blog*, 2019, <<https://openai.com/research/better-language-models>>.
- Radford, Alec, Luke Metz et Soumith Chintala, « Unsupervised representation learning with deep convolutional generative adversarial networks », *ArXiv*, 2016, <<https://arxiv.org/abs/1511.06434>>.
- Sampson, Geoffrey, « Two ideas of creativity » dans Martin Hinton (dir.), *Evidence, Experiment and Argument in Linguistics and Philosophy of Language*, Peter Lang, 2018.
- Shep, Sydney J., « Digital materiality », dans Schreibman, Susan, Ray Siemens et John Unsworth, *A New Companion to Digital Humanities*, Malden, MA et Chichester, West Sussex, UK, Wiley/Blackwell, 2016.
- Vayre, Jean-Sébastien, « Pour une critique plurielle de l'intelligence artificielle », *Revue d'anthropologie des connaissances*, vol. 15, n° 1, 2021, <<https://doi.org/10.4000/rac.19266>>.
- Voices of VR Podcast, « #898 Sundance: Questioning AI of Chomsky trained with 60 years of data in 'Chomsky vs Chomsky' », 12 mars 2020, <<https://voicesofvr.com/898-sundance-questioning-ai-of-chomsky-trained-with-60-years-of-data-in-chomsky-vs-chomsky/>>.
- Weibel, Peter, « The post-media condition », *Arte ConTexto*, n° 6, 2005, p. 11-15, <https://www.ocopy.net/wp-content/uploads/2017/10/weibel-peter_die-postmediale-kondition-englisch-the-post-medial-condition.pdf>.
- Wevers, Melvin et Thomas Smits, « The visual digital turn : Using neural networks to study historical images », *Digital Scholarship in the Humanities*, vol. 35, n° 1, 2019, <<https://doi.org/10.1093/llc/fqy085>>.

3

« DES ARTISTES DE SILICIUM¹ » ? IA ET ROBOTS SUR LE DEVANT DE LA SCÈNE

TECHNICITÉ ET CRÉATIVITÉ DANS LE THÉÂTRE CONTEMPORAIN EN EUROPE

Madelena Gonzalez² et Camille Habault³

L'expression consacrée en français, le « spectacle vivant », semble exclure d'emblée l'intelligence artificielle (IA) et les robots. Pourtant, depuis maintenant près de vingt ans, on a vu émerger, puis se multiplier, des pièces de théâtre qui mettent en scène des robots, mais aussi une tendance croissante à intégrer l'IA aux spectacles. Ce chapitre a pour but d'examiner les rapports entre robotique et spectacle vivant ainsi que les utilisations de robots dans le paysage théâtral contemporain et leurs implications esthétiques et formelles.

Des scénarios catastrophes abondent dans la littérature et la culture contemporaine où l'IA et les robots rivalisent dans la lutte pour la domination ou la suppression de l'humain. En tant qu'art corporel, le théâtre est bien placé pour examiner le concept de l'inhumanité numérique. Effectivement, depuis quelque temps, les utilisations et les représentations d'IA semblent bouleverser les dispositifs performatifs et l'esthétique

-
1. Nous avons emprunté cette expression à Rachid Guerraoui et Lê Nguyễn Hoang (2020, p. 137).
 2. Professeure de littérature anglophone contemporaine à l'Université d'Avignon.
 3. Doctorante à l'Université d'Avignon (2017-2021).

théâtrale contemporaine. Dans ce contexte, la créativité risque-t-elle d'être noyée par un excès de technicité et de produire une déshumanisation de l'expérience théâtrale ? À partir de quelques exemples concrets, cette contribution explore la façon dont la technologie est en train de transformer le théâtre pour le public autant que pour les dramaturges. À l'intérieur de ce nouveau paradigme de création, les robots seraient-ils finalement des artistes de silicium dont la présence témoigne d'un lien fondamental entre l'esthétique et le progrès technologique ?

Depuis ses débuts, le théâtre utilise une forme de technologie pour créer ses illusions que ce soit les poulies et les machines primitives du théâtre antique ou les écrans maintenant omniprésents dans un grand nombre de spectacles. Le mot grec *technè* signifie au départ l'idée de la technique, de l'artisanat, de la production matérielle qui se distingue du concept de *poiësis* qui concernerait plutôt l'acte créatif. L'utilisation croissante des nouvelles technologies dans le contexte de la dramaturgie contemporaine modifie-t-elle le rapport entre ces deux concepts ? Depuis les années 1940 et l'invention du premier ordinateur par Alan Turing, l'essor de la technologie a été exponentiel et a apporté des avancées substantielles dans certains domaines, notamment en médecine où l'on vante les mérites de la robotique. Qu'en est-il de la sphère culturelle et de l'art ?

Il est clair que, dans le domaine du spectacle vivant, les technologies numériques offrent de nouvelles possibilités ; des applications informatiques de traduction instantanée, par exemple, permettent d'« internationaliser » un spectacle bien plus efficacement que des surtitres sommaires et souvent difficilement lisibles. Mais, au-delà de ce type de dispositif fonctionnel dont l'apport est indiscutable, comment la technologie et, plus précisément, l'IA et les robots ont-ils modifié l'esthétique du spectacle vivant ? Puisque le spectacle est « vivant » justement et implique habituellement l'interaction des corps humains dans un espace donné « en direct », l'irruption de l'IA dans l'univers théâtral représente un bouleversement des fondements mêmes du genre. La particularité du théâtre où l'acteur incarne des émotions et les manifestations corporelles de ces émotions sur scène reste-t-elle unique et indépassable ? Peut-on remplacer un acteur par un robot ? L'intégration de l'IA dans un spectacle signifie-t-elle un surplus de *technè* et de technicité aux dépens de l'inspiration de l'artiste dramatique ou, au contraire, ouvre-t-elle un nouvel horizon de créativité artistique ? Comment le paradigme numérique infléchit-il l'esthétique théâtrale ?

Nous allons tenter de répondre à ces questions dans ce chapitre, mais nous devons d'abord situer notre réflexion dans son contexte et décrire les dispositifs théâtraux développés pour intégrer l'IA au spectacle contemporain en Europe. Ensuite, nous examinerons la façon dont la technicité du théâtre contemporain modifie la créativité (la *poïétique*) en la remettant en question, mais en la renouvelant aussi. Pour finir, nous montrerons comment le trouble engendré par l'irruption de l'IA dans le spectacle vivant représente une plus-value esthétique qui dépasse la séparation traditionnelle entre *technè* et *poïésis* et nous incite à réévaluer la place de l'humain, non seulement au théâtre, mais au sein d'une société de plus en plus technologisée.

1. ROBOTS, IA ET CYBORGS : CONTEXTE ET DISPOSITIFS

Le XXI^e siècle étant marqué par la fulgurance parfois vertigineuse de la course au progrès numérique et par les multiples révolutions d'ordre à la fois intime et structurel en découlant, il est évident que le théâtre, toujours bien vivant et reflet de son époque, engloutit ces changements au fur et à mesure, à la fois pour se rendre plus accessible et adapter sa forme, mais aussi pour nous les restituer à sa manière et inciter le spectateur à réfléchir à ces changements sociétaux.

En 2009, Edmond Couchot et Norbert Hilaire considéraient déjà que les arts numériques connaissaient « depuis une dizaine d'années [...] une véritable explosion » (Couchot et Hilaire, 2009, p. 7), faisant remarquer que « presque tous les arts traditionnels empruntent, et de façon croissante, aux technologies numériques, se revivifient et se transforment à leur contact [...] ». Ce ne sont pas uniquement les arts visuels et la musique qui sont sous influence numérique, mais aussi la poésie et la littérature, les arts vivants (théâtre, danse, opéra) » (Couchot et Hilaire, p. 7). Il semble cependant que les utilisations du numérique qu'on pourrait aujourd'hui considérer comme appartenant au courant dominant dans un certain théâtre institutionnalisé se distinguent de la définition de *digital performances*⁴, donnée par

4. D'après Clarisse Badiot : « La traduction française par "performance numérique" est inexacte, le terme "performance" en anglais n'ayant pas la même signification en français. "Arts de la scène et nouvelles technologies" demeure très flou sur le contexte historique : il y a des "nouvelles technologies" à tous les âges de l'histoire du théâtre, et ces "nouvelles" technologies que seraient le numérique et les réseaux ne sont plus si neuves. Quant à "arts de la scène à composantes technologiques", outre la longueur de l'expression, celle-ci met l'accent sur les technologies conçues comme des outils et non comme des environnements de pensée.

Steve Dixon dans son ouvrage de référence : « le terme “*digital performance*” comprend toutes les œuvres des arts de la scène où les technologies informatiques jouent un rôle clé en ce qui concerne le contenu, la technique, l'esthétique ou le résultat final⁵ » (en italiques dans l'original).

Clarisse Badiot commente cette définition de la manière suivante :

Cette définition souligne un point important : ce n'est pas la présence sur scène d'un ordinateur qui permet de définir s'il y a ou non digital performance (sans caricaturer, avec l'essor des régies numériques, on pourrait alors dire qu'aujourd'hui toutes les formes de spectacle vivant sont des digital performances). Les choses sont bien plus complexes que cela. Les technologies numériques peuvent en effet intervenir au cours du processus de création ou encore servir de modèle à l'écriture. Ce n'est donc pas la question de l'outil qui est essentielle mais celle de la pensée à l'œuvre. Cela ne rend pas la définition du corpus aisée mais évite l'écueil d'une vision techniciste des technologies numériques (Badiot, 2013).

On observe un tournant numérique très net en 2020 et en 2021, à la fois dans les lieux de théâtre public, avec deux spectacles très remarquables ces deux dernières saisons, par ailleurs exceptionnelles dans leur déroulement du fait de la crise de la COVID-19 : *Contes et légendes* de Joël Pommerat (metteur en scène adoubé par l'institution s'il en est) et *jeanne_dark* de Marion Siéfert, tous deux interrogeant à leur manière le rôle des technologies numériques dans la vie des protagonistes, avec comme caractéristique, malgré les grandes disparités de traitement entre les deux spectacles, un thème commun : l'adolescence et la construction de l'identité de genre, thématique que l'on retrouve confrontée aux questionnements liés à la robotique dans des écrits théoriques déterminants que nous évoquerons plus loin.

On peut postuler que, contrairement à d'autres technologies numériques plus intégrées dans le quotidien des spectateurs (écrans, ordinateurs, téléphones intelligents, etc.), le robot est difficilement neutre et sa présence visible sur scène ou dans un dispositif performatif ou théâtral, rarement (voire jamais), purement d'ordre pratique.

Le robot étant, de plus, un symbole fort de l'inconscient collectif, façonné depuis des décennies par de multiples œuvres littéraires et

“Représentation numérique” ou “spectacle numérique” ne sont pas très usités et n'évoquent qu'une parcelle du champ étudié » (2013).

5. Il s'agit de la définition de Dixon, reprise et traduite par Clarisse Badiot (2013).

cinématographiques largement populaires, sa présence est troublante, pour le spectateur comme pour le dispositif, même lorsqu'il n'est pas humanoïde.

Ursula K. Le Guin, dans un de ses essais sur la science-fiction, va jusqu'à lui attribuer le statut de « mythologie moderne ». Sur les robots, elle écrit ceci :

On peut considérer le robot comme l'incarnation de la peur de son propre corps qu'éprouve le moi moderne, à la suite de la séparation destructrice du « corps » et de l'« esprit », de l'âge et de la « machine » qu'a opérée la pensée mécaniste après la Renaissance. [...] Ces mythes, ces symboles, ces images ne disparaissent pas, même lorsqu'ils sont placés sous le regard scrutateur de l'intelligence, pas plus qu'un examen éthique, esthétique ou même religieux ne les ferait s'étioier et se dissiper. Au contraire, plus on les regarde, plus leur présence semble se solidifier. Et plus on y pense, plus ils semblent receler de significations (2018, p. 105).

Ce rapport au corps et à sa présence tangible devient plus prégnant encore dans le cas du théâtre, et la dualité corps/esprit encore moins pertinente qu'en littérature. Quel que soit le mode de représentation du robot choisi par le metteur en scène, joué par des actrices dans *Contes et légendes*, par une voix artificielle et un vrai robot non humanoïde dans *Artefact* (2017) de Joris Mathieu ou même dans *Killing Robots* (2020) par des photographies projetées et une reconstitution immobile du robot sur la scène, celui-ci crée une forme de présence nouvelle, et en décalage par rapport à celle, plus habituelle et moins surprenante pour un public occidental, d'un acteur ou d'une actrice jouant un rôle humain.

Les équipes artistiques de Joël Pommerat pour *Contes et légendes* et d'Oriza Hirata pour *Sayônara* semblent attester de la différence de réception des robots présentés sur scène : les spectateurs français de *Contes et légendes*, enfants et adultes, sont, d'après les actrices, très troublés, voire déboussolés, par cette présence, et interrogent la « vraie » nature de ces avatars (sont-ce de vrais robots ?), tandis que les spectateurs japonais de *Sayônara* semblent accueillir la présence du robot avec sérénité. Takenobu Chikaraishi, ingénieur informatique responsable du robot pour *Sayônara*, explique cela par les croyances animistes plus répandues au Japon. Dans *Sayônara*, ce sont Geminoid F, un robot androïde, et Bryerly Long, une actrice, qui se côtoient sur scène, cette dernière témoignant tout de même de la confusion de certains spectateurs (mais peut-être occidentaux ?) qui l'auraient prise, elle, pour le robot.

Cette hybridation des genres humains et robotiques sur scène rappelle également, de manière plus indirecte, le « Manifeste Cyborg » de Donna Haraway, qui utilise un mythe du cyborg⁶ comme métaphore d'une possible libération des normes (binaires) de genre, sexué cette fois. La notion de corps y joue également un rôle important, ainsi que celui de réel et de fiction : « La frontière qui sépare la science-fiction de la réalité sociale n'est qu'illusion d'optique » (2007, p. 30), martèle-t-elle, semblant ainsi refuser toute binarité : réel/fiction, homme/femme, organique/machine, naturel/artificiel :

La fin du XX^e siècle, notre époque, ce temps mythique est arrivé, et nous ne sommes que chimères, hybrides de machines et d'organismes théorisés puis fabriqués ; en bref, des cyborgs. Le cyborg est notre ontologie ; il définit notre politique. Le cyborg est une image condensée de l'imagination et de la réalité matérielle réunies, et cette union structure toute possibilité de transformation historique (p. 31).

Il est troublant de constater que cette « image condensée de l'imagination et de la réalité matérielle réunies » pourrait tout aussi bien décrire le théâtre, et les « nouveaux accouplements fertiles » (p. 31) envisagés pourraient, par exemple, désigner l'arrivée des robots au sein de celui-ci, qui, à leur tour, engendreraient une hybridation du genre et du dispositif théâtral. Le théâtre, tout comme l'humain, serait donc cyborg.

_jeanne_dark, le spectacle de Marion Siéfert montrant une adolescente s'exprimant sur Instagram, et ayant donc lieu à la fois sur le réseau social et dans des lieux tangibles (théâtre, espace public, ou appartements pendant la crise sanitaire maintenant les théâtres fermés), s'il ne montre pas de robot, est peut-être l'illustration parfaite de ce théâtre hybridé et cyborg, mais aussi de l'humain cyborg et façonné par les technologies numériques : l'actrice se filme elle-même avec un téléphone intelligent pendant toute la représentation, tandis que les spectateurs/internautes, qui voudraient la regarder à distance, le font eux aussi à travers leur écran de téléphone.

Contes et légendes et *_jeanne_dark* ont pour thématique centrale la construction de genre chez les adolescents et adolescentes (âge hybride par excellence), et Steve Dixon, dans l'ouvrage cité plus haut, y consacre, et plus précisément consacre à l'esthétique du *camp*, une partie non négligeable dans son chapitre intitulé « Robots ». Le camp est une forme d'expression

6. Selon la définition de Haraway, « le cyborg est un organisme cybernétique, hybride de machine et de vivant, créature de la réalité sociale comme personnage de roman » (2007, p. 30).

issue de la sous-culture gaie et queer, et théorisée par Richard Dyer et Susan Sonntag. Dixon fait le parallèle suivant :

Un certain degré de *camp* semble inhérent à presque toutes les performances robotiques anthropomorphes et zoomorphes que nous avons vues. Les mouvements des robots en sont la clé. Étant donné qu'ils échouent actuellement à reproduire avec exactitude les mouvements humains et animaux, leurs démarches et leurs gestes exagérés produisent la même impression de théâtralité et d'artificialité que l'on retrouve dans le *camp*. L'artiste roboticien Bruce Cannon a déclaré « l'échec de ces machines à transcender leur artificialité constitue leur aspect le plus significatif. Ces objets ne sont pas tant semblables au vivant qu'ils ne sont référentiels du vivant, et la pièce manquante est ce qui m'interpelle ».

Le mouvement robotique imite et exagère, mais ne reproduit jamais l'humain, tout comme les mouvements *camp* imitent et exagèrent, mais n'atteignent jamais parfaitement la féminité. Les fonctionnements robotiques impliquent le même degré de conscience de soi performative que les mouvements *camp*, et souvent leur ressemblent même visuellement. Les mouvements d'un robot anthropomorphe ou d'une personne faisant du *camp* résultent d'une performance dotée d'une conscience de soi ; son codage, son artificialité et sa différence de la norme sont soulignés. Le *camp* devient ainsi central, et même un aspect déterminant de la performance robotique anthropomorphe⁷ (traduction de Camille Habault).

Ces explications résument bien le décalage et le trouble semés par la présence de robots sur scène et reconnaissent le lien étrange entre performance robotique et performance de genre. Ainsi les robots, comme les cyborgs et les femmes, mais aussi comme les acteurs de théâtre, performant

7. « *Some degree of camp seems inherent in almost all performing anthropomorphic and zoomorphic robots we have seen. Their movements are one key to this. Since robots currently fail to mimic human and animal movement accurately, their exaggerated gaits and gestures emphasize the same sense of theatricality and artificiality in movement we find in camp. The artificiality of robot movement mirrors the artificiality of camp. Robotics artist Bruce Cannon has reflected that "the failure of these machines to transcend their artificiality is their most significant aspect. The pieces are not so much lifelike as referential to being, and what is missing is what resonates to me". Robotic movement mimics and exaggerates but never achieves the human, just as camp movements mimic and exaggerate but never achieve womanhood. Robot kinetics involves the same degree of performative self-consciousness as "camping about", and often visually resembles it. When an anthropomorphic robot moves or when a person camps about, it is highly calculated and coded (however ingrained and natural the human camp behavior may have become). Although robots may not yet be self aware, they are quintessentially self-conscious entities, calculating and computing their every move. When a humanoid robot moves, just as when someone camps about, it is a knowing and self-conscious performance; its coding, artificiality, and difference from the norm is emphasized. Camp thus becomes a central, even determining aspect of anthropomorphic robot performance* » (Dixon, 2007, p. 273-274).

et imitent et, si l'on reprend notre postulat de départ pour ce chapitre, on constate que la présence d'un robot n'est jamais neutre.

Le robot et l'IA sont relatifs, sans cesse comparés et mesurés à l'humain, ce qui pour un objet est relativement inhabituel. Par ailleurs, à l'heure actuelle, ils ne peuvent être autonomes que grâce aux efforts d'invention et de programmation de l'humain. Nous savons à présent qu'un robot, même humanoïde, ne peut pas fonctionner comme un humain, que ce soit au niveau de sa motricité, de son énergie ou de son « intelligence ». Pourtant les géants du numérique, tels que Google ou Microsoft, testent régulièrement la *performance* de leurs avancées en matière d'IA en mesurant leurs programmes à des humains dans diverses disciplines (les échecs et le jeu de Go, par exemple), comme l'expliquent Guerraoui et Hoang dans leur chapitre intitulé « Des artistes de silicium » dont s'inspire notre travail (2020, p. 137-152). Ils évoquent notamment le *Portrait d'Edmond de Belamy*, œuvre d'art produite par un algorithme et vendue aux enchères au prix de 432 500 dollars américains en 2018 (p. 140).

Après cette brève présentation du contexte et des dispositifs, nous souhaitons examiner l'interaction entre la technique et l'art créatif afin de voir s'il y a opposition ou dialectique, ce dernier terme étant compris dans le sens de tension entre deux pôles et dépassement de cette opposition.

2. TECHNICITÉ ET POÏÉTIQUE : L'IA AU THÉÂTRE

Au premier abord, l'intégration de l'IA dans un spectacle suggère un surplus de *technè* et de technicité aux dépens de l'inspiration de l'artiste dramatique qui modifierait substantiellement l'acte créatif, sa *poiétique*. Mais cette évolution ne pourrait-elle pas aussi ouvrir un nouvel horizon de créativité artistique ?

Dans la lignée d'une réinterprétation du terme *technè*, comme on le trouve dans les travaux de Henry Staten par exemple, nous allons analyser la (fausse ?) dichotomie entre *technè* et *poiésis*. Comme il l'explique dans « The Wrong Turn in Aesthetics », une relecture d'Aristote et de Platon, qui écarte le culte du génie, développé par Kant et le courant romantique, permet de réévaluer ces concepts :

Technè, au contraire, est [...] la partie « matérielle » de la technologie qui est déjà nécessaire pour créer la forme de l'œuvre : la partie qui fait appel aux outils, aux matériaux, à la manipulation habile de ces outils afin de

donner une forme aux matériaux, et aussi au savoir qui entoure tout ce processus [...].

L'opinion moderne n'a pas tendance à reconnaître que les œuvres d'art sont le produit de la *technè*; elle a tendance à privilégier des conceptions « romantiques » et « humanistes » qui soutiennent que, pour créer une œuvre d'art qui mérite ce nom (le « beau » ou l' « inspiré », comme l'appelle Kant), le facteur essentiel, ce qui fait que c'est *vraiment de l'art*, doit être quelque chose de plus que la *technè*, quelque chose qui demande une intervention spéciale, et indéfinissable de la part de l'artiste⁸ (traduction : Madelena Gonzalez).

L'hypothèse de Staten est la suivante : l'artisanat et l'acte créatif sont inséparables. D'après lui, Aristote ne craignait pas la machine ni la technicité ; pour lui, la main était une prothèse naturelle et la nature humaine était essentiellement d'ordre « prosthétique » (Staten, 2012, p. 50). Ces réflexions rappellent les propos échangés par Jean-Luc Nancy et Benoît Goetz dans un entretien sur la technique et l'esthétique :

Benoît Goetz : « Leroi Gourhan, en particulier, dit quelque chose de très proche de tes propres considérations lorsqu'il suggère que l'on pense très mal à la technique quand on la voit comme quelque chose d'extérieur à l'homme, qui lui tomberait dessus de l'extérieur, alors qu'elle est son prolongement, l'extériorisation de son squelette, de son système nerveux, de son imaginaire ».

Jean-Luc Nancy : « Je crois même qu'il faudrait dire encore un peu plus qu'« extériorisation » ou « prolongement »..., il faudrait dire que c'est l'homme lui-même qui est l'animal technique » (Nancy, 1999, para. 3 et 4).

Examinons à présent quelques exemples concrets de l'interaction entre *technè* et *poïésis* quand il s'agit de mettre en scène des robots. Oracle de Susanne Kennedy (2020) propose un dispositif déambulatoire au terme duquel le spectateur, seul (mesures sanitaires obligent), sera confronté à un robot oracle à qui il pourra poser des questions sur le futur. Il s'agit donc d'un robot conversationnel (*chatbot*), dont un certain nombre de réponses sont probablement programmées à l'avance, mais qui réagit tout de même en direct, en fonction des interrogations de son interlocuteur, et qui en

8. « *Technè, by contrast, is [...] the "material" part of the technology that is already involved in the shaping of the work: the part involving tools, materials, the skillful wielding of these tools to shape the materials, and the knowledge in and around all of this [...]. Modern opinion has not been much inclined to the notion that artworks are made by technè; it has tended rather to "Romantic" and "humanistic" conceptions which hold that, in order to make a work of art worthy of the name ("beautiful" or "inspired" art, Kant calls it), the essential factor, what really makes it art, must be something beyond technè, something that requires a special, indefinable intervention on the part of the artist* » (Staten, 2011, p. 226).

quelque sorte génère le ou les textes de l'œuvre en temps réel, en collaboration avec le public. Spectateur et IA deviennent alors co-auteurs et co-acteurs, tandis que le spectateur endosse le rôle à la fois du spectateur (puisque l'interaction n'a lieu devant personne d'autre), d'un visiteur, d'un acteur et d'un auteur du spectacle. Le rôle de la metteuse en scène s'en trouve lui aussi déplacé. Cette œuvre a pourtant bel et bien été montrée dans un théâtre l'été dernier (au Kammerspiele à Munich), et Susanne Kennedy se définit bien comme metteuse en scène de théâtre.

Cet exemple montre clairement que la technologie a déplacé et modifié le rôle du créateur. Le spectacle est le produit de plusieurs instances créatives et de l'interaction entre ces instances. Le dramaturge est autant artisan que poète et la technicité est indissociable de l'acte créatif. On pourrait même aller plus loin et dire que les nouvelles technologies démocratisent la *technè* qui n'est plus réservée à l'artisan-poète, mais beaucoup plus largement disponible et, surtout, partagée. Créer n'est plus un acte individuel et inspiré mais un acte collaboratif à partir de dispositifs technologiques sophistiqués, associés à l'imaginaire collectif.

Cette impulsion collaborative induite par la technologie et cet éloignement de la vision traditionnelle de l'artiste inspiré viennent renforcer une évolution déjà présente dans le spectacle vivant. Également transformés par la présence et la représentation d'IA, textes dramatiques et écritures théâtrales semblent prendre plus généralement la forme d'écritures de plateau, tendance que l'on constatait déjà d'une manière générale dans le théâtre contemporain, mais qui semble se systématiser dans les « spectacles numériques », particulièrement dans les spectacles incluant robots et IA. En effet, la plupart d'entre eux ne donnent pas lieu à une publication de leur texte.

C'est le cas, par exemple, de la dernière pièce de Joël Pommerat, *Contes et légendes* (2019), évoquée plus haut et qui a pour thématique centrale les robots. Chaque robot est joué par une actrice, et le spectacle n'inclut donc pas d'IA réelle. Le dispositif du spectacle est frontal et classique, et le récit en est linéaire. C'est pourtant un des rares spectacles de Joël Pommerat, qui se définit comme « écrivain de spectacles », dont le texte, malgré le succès du spectacle, ne donnera pas lieu à une publication. Peut-on alors supposer que la seule thématique de l'IA dans le spectacle vivant semble s'opposer à la pratique du texte dramatique écrit et publié à l'adresse du public ? Si c'est le cas, cela transforme de manière importante le théâtre en tant qu'objet à la fois littéraire et performatif et c'est le performatif qui prend le dessus.

D'autres spectacles, il est vrai, rendent la publication de leur texte compliquée, étant donné leur dispositif. La simple utilisation en direct sur scène d'une IA sous forme de robot conversationnel (*chatbot*), c'est-à-dire d'un outil conversationnel réagissant à la voix de l'utilisateur, rend la représentation imprévisible, puisqu'on ne sait pas en général par avance ce que va dire le robot conversationnel. Mais cela va parfois encore plus loin, puisque l'IA et les technologies numériques ont tendance à faciliter et à encourager les interactions du public avec l'œuvre, redéfinissant parfois entièrement sa place pour en faire une partie intégrante du dispositif. C'est le cas de *Not Electronic City* adapté par Emmanuel Lagarrigue à partir d'une pièce de Falk Richter (2017). Il s'agit d'une installation proposant une pièce de théâtre jouée simultanément sur plusieurs écrans, donc fragmentée. C'est au spectateur de reconstituer la pièce à sa guise, dans l'ordre qu'il veut, en s'attardant ou non sur les écrans qu'il choisit : dans ce contexte, le dispositif lui-même est le robot, un théâtre machine qui englutit son public.

Dans *Artefact* de Joris Mathieu, qu'il définit lui-même comme un spectacle/installation sans comédiens, les spectateurs/visiteurs se voient prêter un casque audio, dans lequel ils entendent par la suite une voix synthétique, celle de l'IA, également personnage principal de cette œuvre, qui les guide à travers le dispositif tout en leur racontant son histoire. Les spectateurs sont divisés en trois groupes, reconnaissables par la couleur du voyant lumineux allumé sur leurs casques. Chaque groupe obéit docilement en temps réel en fonction des instructions données dans le casque par le personnage de l'IA, docilité dont « la voix » dans les casques n'hésite pas à se moquer, se demandant si c'est plutôt elle ou les humains qui agissent le plus comme des machines et sont le plus obéissants. Les spectateurs d'*Artefact* font donc partie intégrante du dispositif et du type d'esthétique qui y est convoqué. Par leurs actions, ils matérialisent eux-mêmes la présence et l'intelligence de l'IA, qui dans l'histoire veut d'ailleurs créer un spectacle théâtral : les spectateurs humains deviennent ses marionnettes. En revanche, la forme du spectacle dans sa dimension métathéâtrale et réflexive, deux caractéristiques qui appartiennent à l'esthétique post-moderniste⁹, les rend conscients de ce qui leur arrive plutôt que passifs, donc cocréateurs en quelque sorte.

9. Les théoriciens de l'esthétique postmoderniste, comme Linda Hutcheon (1989) et Charles Jencks (1992) attirent l'attention sur l'aspect autoréflexif de l'art postmoderne.

Il semblerait que l'intégration des robots permet de problématiser notre conception du théâtre et de la place de l'humain dans l'art. Les robots sont-ils des artistes de silicium source d'une esthétique collaborative qui pourra renouveler l'art dramatique ou matérialisent-ils sur scène l'inhumanité numérique qui, à l'époque actuelle, fait abstraction du corps humain pour imposer un rythme infernal à notre quotidien ? Dans notre dernière partie, nous essaierons de déterminer si l'utilisation de nouvelles technologies au théâtre signale un changement de paradigme esthétique et, si c'est le cas, quelles en seraient les conséquences éventuelles sur le plan humain.

3. « ARTISTES DE SILICIUM », ROBOTS PERFORMANTS ET PERFORMATIFS

L'origine grecque du mot esthétique αἰσθητικός (*aisthetikos*), repose sur la notion de perception et fait appel à nos cinq sens. Le théâtre, art corporel par excellence, a un rôle important à jouer dans la réflexion urgente à mener sur le sujet de l'incarnation de l'humain dans les technologies numériques et vice versa, et des répercussions des IA sur nos sociétés et nos imaginaires collectifs. Le théâtre est aussi la manifestation visible de nos émotions et de nos pensées et, en tant que tel, il peut faire avancer la réflexion sur l'interaction humain-IA. Dans cette dernière partie, nous essaierons de démontrer au moyen de quelques exemples de spectacles contemporains que l'esthétique, composante essentielle de l'existence humaine et besoin humain fondamental, est inséparable du progrès technologique.

Dans un documentaire diffusé sur Arte en 2020 au sujet de l'IA, un ingénieur IA (de Facebook) évoque la possibilité que, selon lui, les IA vont permettre aux humains (qui n'auront plus à travailler) de se consacrer plus largement à l'art, seule activité irremplaçable par des machines ou l'IA, puisqu'elle part de l'émotion et a pour but d'en susciter. Cela est un point de vue assez répandu qui rappelle l'intrigue de *R.U.R.* de Karel Čapek, pièce de théâtre de science-fiction écrite en 1920 où les robots existent pour soulager les humains des tâches manuelles et ingrates et leur permettre de se consacrer aux choses plus nobles de l'esprit. On y retrouve la dichotomie corps/esprit et la dévaluation du corps par rapport à l'esprit : le robot se voit placé sur le même niveau de valeur que le corps humain.

R.U.R. se solde par l'apocalypse car les robots développent sentiments et émotions : un corps ne peut pas n'être qu'un corps séparé de tout le reste.

Aussi l'ingénieur a peut-être tort : un robot peut très bien susciter une émotion, particulièrement au théâtre, art de l'imitation et de

l'interprétation par excellence : les robots, eux aussi, imitent, et leur décalage, qu'il soit infime ou flagrant, avec les comportements humains, est peut-être précisément ce qui crée l'émotion.

La réalisatrice et auteure allemande Hito Steyerl nous propose dans *Hell Yeah We Fuck Die* (2016) une

installation qui aborde avec humour l'horizon menaçant d'intelligences artificielles amenées à remplacer l'humanité, les images vidéo s'accrochent à des structures tubulaires assorties de barricades de chantier, tandis que le public lui-même prend place sur les mots du titre. Ces méandres métalliques peuvent aussi être compris comme l'inconscient des machines : le réseau numérique qui engendre des simulacres. Les vidéos montrent l'ultime combat des êtres humains contre leurs créatures : des images trouvées de laboratoires testant la stabilité des robots humanoïdes qui ont déjà envahi l'espace de l'exposition (Centre Pompidou, 2016).

Les vidéos montrent le côté ridicule, « vulnérable », de vrais robots humanoïdes (on voit bien qu'ils ne sont pas prêts à nous remplacer), et des humains qui les poussent et leur donnent des coups pour les faire vaciller. Le spectacle relève aussi de l'esthétique particulière du bricolage, avec des robots/formes humanoïdes très rudimentaires au sol, des vidéos d'une qualité moyenne, non filmées au départ dans un intérêt esthétique. Un tel dispositif permet néanmoins d'aborder (et peut-être d'exorciser) des thèmes profonds concernant notre angoisse par rapport à l'avenir et la place de l'humain dans un univers technologisé.

Il est clair que les technologies nouvelles, et plus particulièrement les robots, sèment le trouble au théâtre, que ce soit par le rire, la peur ou l'étrangeté, et c'est sans doute précisément dans ce trouble que se situe la plus-value artistique de la rencontre entre l'IA et la scène. *Killing Robots* de Linda Blanchet est un spectacle-enquête canadien, joué au Théâtre de Paris-Villette en 2020, qui raconte l'histoire vraie de HitchBot, un robot auto-stoppeur, créé par deux laboratoires canadiens pour étudier le lien entre l'homme et la machine et qui finit par être « assassiné ». C'est grâce à des photos un peu floues, aléatoires et mal cadrées, que le robot crée sa propre esthétique du trouble, amplifiée par la mise en scène avec musique, lumières inquiétantes, thématiques du meurtre, vandalisme et enquête policière. Le spectacle se présente comme du théâtre documentaire se racontant beaucoup à travers les photographies prises par le robot auto-stoppeur. Une enquête se déroule pour essayer de comprendre le « meurtre » du robot. Le public est encouragé à se demander pourquoi on s'en est pris à

ce robot dont le côté enfantin le rend sympathique et, du coup, la question des émotions, de l'empathie et de ce qui constitue l'humain est posée dans toute sa complexité.

Certains robots, pourtant, ne sont *a priori* pas pensés pour interagir avec l'humain ni pour lui ressembler, mais pour accomplir efficacement une tâche précise ; par exemple, les bras robotiques qui sont utilisés depuis plusieurs décennies maintenant, dans l'industrie automobile surtout. Mais, lorsqu'ils sont placés sur scène et arrachés à leur utilité première, leur dimension performative redevient manifeste. C'est le chemin qu'a suivi dès 2008 le chorégraphe Aurélien Bory avec son spectacle *Sans objet* où la confrontation à la fois drôle et inquiétante entre les danseurs/acteurs et l'énorme robot qui domine la scène interroge les frontières de l'humain, comme l'explique Christophe Lemaire :

Sans objet, que j'ai vu à la création à Vidy-Lausanne, est à la fois drôle et inquiétant : le corps humain y est d'abord performant et le robot sensible, ce qui rapproche d'un vocabulaire à la fois burlesque et fantastique (ou poétique), puis l'ensemble glisse insensiblement vers une déshumanisation effrayante et la machine finit par faire la démonstration concrète de son pouvoir et de sa puissance physique (2010).

L'aller-retour entre la robotisation de l'humain et l'humanisation des robots, une sorte d'inquiétante étrangeté dans les deux sens, incite à repenser le rapport entre l'homme et la technique.

Plus récemment, le metteur en scène Joris Mathieu montrait lui aussi un bras robotique dans son spectacle/installation *Artefact*, qui raconte l'histoire d'une IA voulant faire du théâtre après l'apocalypse, et propose donc au spectateur de découvrir un dispositif composé de machines lui donnant l'illusion de générer une œuvre théâtrale de manière autonome et d'y pénétrer. Ainsi les machines ne sont pas simplement présentées ou représentées sur une scène de théâtre, elles côtoient directement le public, qui déambule à travers elles, guidé par une voix artificielle dans un casque distribué à l'entrée. L'esthétique de ce spectacle est inquiétante ; une tête de mort, une machine à fumée, une voix grave et menaçante qui alterne avec l'intonation mécanique et stéréotypée de la voix synthétique contribuent à créer une ambiance de fin de monde. Les mouvements du Bras Kuka sont impressionnants et étonnants de fluidité et produisent une sorte de danse très harmonieuse et troublante de la part d'un bras robotique industriel qui n'est pas sans rappeler l'esthétique de *Sans objet*.

Peu après les débuts des représentations d'*Artefact* et étant donné l'absence dans l'œuvre d'humain autre que le public, l'équipe artistique a ressenti la nécessité de mettre en place un « échange humain » après chaque représentation, c'est-à-dire un échange avec un ou plusieurs membres de la compagnie afin que les spectateurs/visiteurs, mais aussi les artistes, proposant l'expérience, aient l'occasion de mettre des mots sur ce qu'ils venaient de vivre et de voir. « Bienvenue au théâtre ou ce qu'il en reste », annonce Philippe Charreron, un des comédiens de la compagnie pour ouvrir l'échange humain du 21 juillet 2018 dans le Festival OFF d'Avignon. Toujours au cours de cet échange, à propos des difficultés du travail avec les robots, il témoigne :

La machine ne se fatigue pas, par rapport à nous. Ça rend les interactions bien plus compliquées que lorsque c'est un comédien au plateau. C'est usant. Avec l'équipe de comédiens, on était en tournée sur un autre spectacle, donc nos camarades créaient à Lyon, et on revenait régulièrement voir où ça en était, on partageait un peu ce qui se passait et plusieurs fois, quand on est revenus, on s'est dit qu'il fallait les sortir de cet enfer des machines. Ils étaient devenus tous gris, à force de s'acharner sur les imprimantes 3D et d'essayer de trouver tout ça... (propos recueillis par Camille Habault, 21 juillet 2018).

Cette réaction négative semble indiquer qu'il y a bel et bien changement de paradigme. L'automatisation grandissante de la société, les progrès déjà impressionnants en IA, les ambitions transhumanistes de la Silicon Valley interrogent et inquiètent beaucoup de monde. Le philosophe Jean-Michel Besnier va jusqu'à dire que « l'IA est en train de tuer l'humain en nous » (Alexandre et Besnier, 2016, p. 82). Il est certain que de telles projections influent sur notre discours et conditionnent les représentations des artistes.

CONCLUSION

De toute évidence, l'entrechoquement entre technicité et créativité au théâtre est douloureux. Mais n'est-il pas, d'après les exemples étudiés, aussi source d'innovations, d'expérimentations et d'interrogations fertiles ? L'irruption de robots et de l'IA dans le spectacle vivant force le théâtre à réfléchir sur sa forme et les frontières entretenues avec d'autres genres et disciplines artistiques, la présence de robot ajoute une couche de simulateur/imitation/trouble au théâtre, qui en est déjà le miroir. Le robot ne permet-il pas également de réexaminer la place de l'humain dans le spectacle vivant et dans la société, tout en renouvelant l'esthétique théâtrale ?

Le théâtre est le lieu du corps, mais aussi de l'objet symbolique, tous deux interrogés par le robot, comme le sont les dualités corps/esprit, mise en scène/texte, scène/public, art/technique qui se voient sans cesse réévaluées. En tant qu'art vivant, précisément, le théâtre ne peut rester figé en dehors de son temps.

Guerraoui et Hoang définissent la notion de créativité de la manière suivante : « On l'associe souvent avec la capacité de générer une information, qu'il s'agisse d'une idée ou d'une œuvre, qui soit à la fois surprenante et fascinante. Ces deux propriétés semblent être des clés pour permettre à l'idée ou à l'œuvre de nous interpeller » (p. 141-142). Le théâtre contemporain qui met en scène et utilise l'IA, les robots et les nouvelles technologies témoigne clairement de cette dynamique qui réunit *technè* et *poiétique* dans une dialectique angoissante, certes, par sa mise en question de la place de l'être humain dans l'acte créatif, mais féconde, aussi, grâce à la plus-value du trouble qu'elle apporte sur le plan esthétique. *In fine*, ne nous offre-t-elle pas la possibilité de dépasser la séparation traditionnelle entre technicité et créativité ?

RÉFÉRENCES

- Alexandre, Laurent et Jean-Michel Besnier, *Les robots font-ils l'amour ? Le transhumanisme en 12 questions*, Malakoff, Dunod, 2016.
- Badiot, Clarisse, « Que sont les "digital performances" », *Leonardo Olats/Leonardo online*, février 2013, <http://archive.olats.org/livresetudes/basiques/artstechnosnumerique/1_basiquesATN.php>.
- Blanchet, Linda et Cie Hanna R., *Killing Robots*, Théâtre Paris Villette, 2020.
- Bory, Aurélien, *Sans objet*, Théâtre national de Toulouse Midi-Pyrénées, 2009.
- Čapek, Karel, *RUR : Rossum's Universal Robots : drame collectif en un prologue de comédie en trois actes*, trad. par Brigitte Munier, Paris, La Différence, 2011.
- Centre Pompidou, *Hell Yeah We Fuck Die*, Catalogue de l'installation, 2016.
- Couchot, Edmond et Norbert Hilaire, *L'art numérique : comment la technologie vient au monde de l'art*, Paris, Flammarion, 2009.
- Dixon, Steve, *Digital Performance : A History of New Media in Theatre*, Cambridge, MIT Press, 2015.
- Guerraoui, Rachid et Lê Nguyễn Hoang, *Turing à la plage : l'intelligence artificielle dans un transat*, Paris, Dunod, 2020.
- Haraway, Donna, « Manifeste Cyborg », traduit par Marie-Hélène Dumas, Charlotte Gould et Nathalie Magnan, dans Laurence Allard, Delphine Gardey et Nathalie Magnan (dir.), *Manifeste cyborg et autres essais : sciences, fictions, féminismes*, Paris, Exils, 2007, p. 29-106.
- Hirata, Oriza. *Les trois sœurs (version androïde)* suivi de *Sayônara (version 2)*, trad. par Mathieu Capel et Hirotoishi Ogashiwa, Besançon, Les Solitaires intempestifs, 2014.

- Hutcheon, Linda, *The Politics of Postmodernism*, London and New York, Routledge, 1989.
- Jencks, Charles, « The Post-Modern Agenda » dans C. Jencks (dir.), *The Post-Modern Reader*, London, Academy Editions, 1992, p. 10-39.
- Kennedy, Susanne, *Oracle*, Kammerspiele, Munich, 2020.
- Lagarigue, Emmanuel, *Not Electronic City* (d'après Falk Richter), Paris, Palais de Tokyo, 2017.
- Le Guin, Ursula K. *Le langage de la nuit*, traduit par Francis Guévremont, Paris, Le Livre de poche, 2018.
- Lemaire, Christophe, « Interview d'Aurélien Bory par Christophe Lemaire », janvier 2010, <<https://www.theatre-contemporain.net/spectacles/Sans-objet/ensavoirplus/>>.
- Mathieu, Joris, *Artefact*. Théâtre Nouvelle Génération, Lyon, 2017.
- Nancy, Jean-Luc, Entretien avec Benoît Goetz, « Techniques du présent », *Le Portique*, vol. 3, 1999, <<https://journals.openedition.org/leportique/309>>.
- Pommerat, Joël, *Contes et légendes*, CDN Nanterre Amandiers, 2019.
- Schei, Tonje Hessen, « IHuman – L'intelligence artificielle et nous », Norvège 2019, diffusé sur Arte le 21 avril, 2020.
- Siéfert, Marion, *_jeanne_dark_*, La Commune, Aubervilliers, 2020.
- Staten, Henry, « The Wrong Turn of Aesthetics », dans Jane Eliot and Derek Attridge (dir.), *Theory after « Theory »*, Abingdon, Routledge, 2011, p. 223-237.
- Staten, Henry, « The Origin of the Work of Art in Material Practice », *New Literary History*, vol. 43, n° 1, 2012, p. 43-64.

4

ÉMERGENCE DES DISPOSITIFS D'AUTOMATISATION DANS LE JEU DE SOCIÉTÉ

LE RÔLE DE L'INTELLIGENCE ARTIFICIELLE

Samuel Francblu¹

Historiquement, les gens ont toujours eu des idées naïves au sujet des qualités qui, une fois mécanisées, constitueraient indéniablement l'intelligence. Il semble parfois qu'au lieu de produire un résultat qui, de l'avis de tous, est l'intelligence réelle, chaque nouvelle étape vers l'IA dévoile uniquement ce que l'intelligence *n'est pas*.

Hofstadter et Denett, 1987.

À L'ATTENTION DU LECTEUR

Afin d'alléger la lecture, nous utiliserons dans ces pages, indifféremment et comme synonymes, les termes « jeu de société », « jeu de plateau² » ou « jeu de table » pour désigner les jeux de support matériel, « physique » ou encore « analogique » – par opposition au support *numérique* du jeu vidéo³ (Wake, 2019) –, en dépit des différentes taxinomies qui peuvent être utilisées par divers acteurs⁴.

1. Doctorant-chercheur en sciences de l'information et de la communication à l'Université de Lorraine.
2. Le terme *jeu de plateau*, équivalent littéral de l'anglais *boardgame* ou de l'allemand *Brettspiel*, est souvent utilisé par les joueurs eux-mêmes du fait de l'ambiguïté du terme *jeu de société* : un nombre conséquent de ces jeux se pratiquent en effet aussi en « solo » (Leorke, 2018).
3. En anglais américain *analog games*.
4. La fiche Wikipédia nous semble à ce sujet particulièrement révélatrice. Wikipédia, « Jeu de société », <https://fr.wikipedia.org/wiki/Jeu_de_soci%C3%A9t%C3%A9> (consultée le 3 février 2022).

cadres de sociabilité spécifiques et repose sur des mécanismes procéduraux inspirés du jeu vidéo.

Si l'importation de dispositifs numériques dans le matériel des jeux de société (le « jeu hybride ») a pu être traitée par ailleurs (voir Chouffot, 2019), tout comme la numérisation des jeux de plateau classiques (par exemple : Bustros, 2009, ou Bard *et al.*, 2019). Il s'agit plutôt ici de comprendre le potentiel, en matière de création, de médiation, de vulgarisation, de l'application au jeu de société de mécanismes inspirés des savoir-faire numériques et en particulier de l'IA. Ce chapitre rend compte d'une recherche qui articule une enquête de terrain au sujet des usages avec une analyse d'inspiration sémio-communicationnelle des mécanismes ludiques mis en œuvre.

Après avoir brièvement présenté le champ du jeu de société et le contexte d'émergence de ses formes contemporaines, nous proposerons une méthode pour analyser cette pratique grâce à des notions inspirées des recherches sur le jeu vidéo.

Notre deuxième partie propose au lecteur un panorama des interactions récentes entre les deux traditions (jeu numérique/jeu analogique), puis l'analyse d'un corpus de procédures d'automatisation couramment mise en œuvre dans le jeu de société. Ce cadre théorique nous permettra en retour de saisir les limites des dispositifs « analogiques » inspirés de l'IA.

Nous tenterons alors d'évaluer la pertinence de l'application de ces mécanismes dans le champ de la création ludique « analogique », puis en matière d'usage social. Il s'agit donc d'envisager l'IA dans le jeu de société, d'une part comme objet culturel, d'autre part comme pratique sociale. Enfin, nous chercherons à évaluer le potentiel du jeu en matière de vulgarisation de l'IA, c'est-à-dire à aborder le jeu de société comme vecteur de connaissance.

1. CONTEXTE ET ENJEUX DE L'AUTOMATISATION DANS LE JEU DE SOCIÉTÉ

1.1 Émergence des dispositifs d'automatisation dans le champ du jeu de société

La démocratisation du jeu de société moderne⁵ laisse à croire que l'activité ludique est en passe de « s'imposer de façon durable dans les pratiques

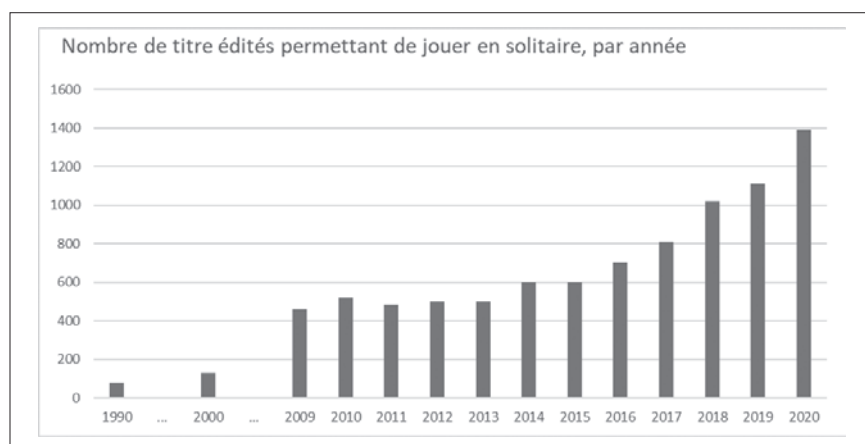
5. Ce jeu dont la « diffusion commerciale n'est pas un caractère extérieur au jeu mais intrinsèquement lié à ce qu'il est », et dont les « règles ne sont pas transmissibles isolément du matériel complexe

culturelles» (Berry, 2018). Bien qu'elle prenne une place grandissante dans les pratiques de loisirs ainsi que dans l'industrie culturelle (Brougère, 2003 ; Dauphagne, 2010, 2011), le monde de la recherche l'aborde encore rarement dans le cadre des représentations culturelles qui l'étaient et qu'elle produit.

Notre approche de ce secteur cherche à ouvrir la boîte noire du jeu de société comme objet culturel en nous appuyant sur l'analyse de quelques titres contemporains et de leur utilisation par une « communauté de pratique » (Brody, 2015), car les nouveaux mécanismes narratifs répondent à des enjeux qui concernent autant la conception que l'usage.

Avec l'âge et l'expérience, l'activité ludique prend en effet place dans une constellation de pratiques où se développe une sociabilité basée sur des réseaux d'affinités. Pour ce faire, des dispositifs intégrés à un nombre croissant de jeux (figure 4.2) permettent un fonctionnement *joueur(s) contre jeu* sans nécessiter de concertation préalable (Francblu, 2020).

FIGURE 4.2
**Nombre de titres édités par année, intégrant
au moins un mode automatisé**



Champ : 85 000 titres répertoriés

Source : l'auteur, d'après des données issues de BoardGameGeek <www.boardgamegeek.com>.

dont elles donnent les conditions de manipulation» (Brougère, 1979).

Depuis une dizaine d'années (Chouffot, 2019), les designers cherchent à construire des instances de régulation ou de narration qui recouvrent des systèmes de génération des univers fictifs – permettant, par des mécanismes de narration procédurale, de créer une intrigue nouvelle à chaque partie – et des systèmes d'activation des adversaires ou même des *joueurs absents*.

Ces systèmes de jeu sont décrits comme des IA par les éditeurs et par le public, en ce qu'ils remplacent le *maître de jeu* du traditionnel jeu de rôles. Or dans un champ où une certaine porosité⁶ s'observe entre le monde des *designers* et le monde des consommateurs, les pratiques ludiques dites « de niche » semblent constituer un modèle innovant pour tout le secteur de la création (Zabban, Piñeros et Roux, 2021).

1.2 Jeu numérique et analogique : un panorama

Si, au début du jeu vidéo, la plupart proposaient de simples affrontements, les univers se sont complexifiés, impliquant une certaine *autonomie narrative*. L'implémentation de méthodes d'IA représente alors dans le jeu vidéo l'occasion « d'améliorer le comportement des personnages non joueurs (PNJ) ou encore [de] générer du contenu de façon automatique », constituant un ensemble de réponses comparables à un ensemble de problèmes identiques (Collet, 2018).

Les méthodes de l'IA *ad hoc* sont donc utilisées pour définir les actions des agents (au sens de Murray, 1997). Ces « entités agissantes » peuvent être aussi bien des éléments de l'environnement que des PNJ ou « personnages non joueurs ». L'univers vidéoludique a par ailleurs hérité du jeu de rôle ce terme qui désigne tous les personnages de la fiction, adversaires ou alliés, qui ne sont pas incarnés par les joueurs, mais mis en scène par le *maître de jeu* ou, dans le cas du jeu vidéo, par la machine.

Le développeur définit manuellement des états et des seuils qui peuvent faire passer l'agent d'un état à un autre. Les « arbres de comportement » ont une structuration proche, mais reposent sur des actions et non des états. La transition entre les actions est représentée par une arborescence qui permet de régir des actions plus complexes.

6. D'une certaine manière en effet, les joueurs d'aujourd'hui sont les concepteurs de demain.

1.3 Des procédures inspirées du numérique dans le jeu physique

Dès lors, l'analogie peut être faite avec les règles de constitution du plateau de jeu, ou avec les procédures de construction de la trame narrative dans le jeu de société (voir figure 4.3 avec l'exemple de *Fallout*, le jeu de plateau).

FIGURE 4.3
Construction procédurale de la trame narrative dans *Fallout*



Source : photographie de l'auteur.

Dans un contexte de succès des jeux coopératifs, supposant l'affrontement du groupe de joueurs non contre des adversaires, mais des *événements*, des algorithmes sont imaginés par les concepteurs pour générer les réactions de l'environnement dans les jeux de plateau, comme le mécanisme d'inondation de *L'Île interdite* (2010) ou celui de la contagion planétaire dans *Pandemic* (2009) (voir figure 4.4).

Répondant à la nécessité de s'adapter au cadre temporel et organisationnel de la pratique, des *tutoriels* sont rédigés par les concepteurs. Ce

sont des suites de procédures narratives qui permettent de mettre en place la partie *en temps réel*, d'apprendre des règles complexes tout en jouant (par exemple : *Andor*, 2012).

FIGURE 4.4
Progression automatisée de la menace planétaire dans *Pandemic*



Source : photographie de l'auteur.

D'un autre côté, les *jeux d'évasion* sur table et autres jeux d'énigmes constituent des expériences uniques, dans le sens où l'intrigue est dévoilée par la partie, avec un investissement minimal en préparation. Souvent, ces dispositifs s'appuient sur la construction « d'intrigues émergentes » (Caïra, 2014) ou de « foyers de tension narrative » (Marti et Baroni, 2014), comme dans les titres primés lors des festivals internationaux de Cannes et d'Essen : *Time Stories* (2015), *Unlock!* (2017), *Exit* (2017) ou encore *Détective* (2018).

L'expérience de jeu implique parfois la transformation du matériel (l'héritage laissé par une session sur les sessions suivantes, ou *legacy*) selon une trame semi-aléatoire, parfois au cours d'une *campagne* qui permet de décliner un scénario sur plusieurs parties. *Mice and Mystics* (2012), *The 7th Continent*

(2017), *Folklore the Affliction* (2017), *Détective* (2018), *Voyages en Terre du Milieu* (2019), *Comanauts* (2019), *Etherfields* (2020) ou *Tainted Grail* (2020) construisent ainsi leur intrigue au moyen d'une arborescence narrative.

Les jeux les mieux « cotés⁷ » du secteur, *Mage Knight* (2011), *Les Demeures de l'Épouvante* (2011), *Lords of Hellas* (2018), *Historia* (2014), *Outlive* (2016), *Gloomhaven* (2017) ou encore *Zombicide* (2012) offrent des systèmes d'activation de personnages et d'adversaires fictifs qui permettent aux joueurs humains de coopérer, d'affronter et d'interagir avec des PNJ (voir pour illustration les figures 4.5 et 4.6).

FIGURE 4.5
Déplacements des Personnages Non-Joueurs
dans *Xia – Legends of a Drift System*



Source : photographie de l'auteur.

7. D'après le site BoardGameGeek, <www.boardgamegeek.com>, référence en la matière avec plus d'un million d'abonnés.

FIGURE 4.6

Cartes de réactions des ennemis automatiques dans *Okko Chronicles*



Source : photographie de l'auteur.

1.4 Enjeux liés à l'IA « analogique »

L'objet de cette recherche peut donc se définir comme l'analyse de ces mécanismes appliqués au jeu de société. Ici, c'est une mécanique de

papier et de carton qui fait figure d'« intelligence artificielle », qui propose des arbres de décision capables de prendre en compte le contexte et les actions des joueurs : le texte et les règles du tirage des cartes remplacent simplement les puces et les lignes de code.

Au sein d'un large éventail de pratiques ludiques, une « communauté d'acteurs » conçoit et utilise un certain nombre de dispositifs qui « irriguent » tout le champ. Notre approche du jeu comme « artefact socio-culturel » (Simonian *et al.*, 2016), ou contenu technique répondant à un certain usage social, nous amène à examiner le rôle et la place de l'IA dans le jeu analogique selon un triple point de vue.

Quels sont les points de comparaison possibles avec l'IA utilisée dans le jeu vidéo ? Quels apports de la recherche en IA peut-on appliquer aujourd'hui au jeu de société ?

De plus, si le design contemporain répond à une réorganisation des cadres de la pratique, à quel besoin l'automatisation des systèmes de jeu peut-elle apporter une réponse ? En quoi ces systèmes ont-ils des répercussions sur la place du jeu dans l'espace social des pratiques ? Il s'agira en définitive de vérifier si les mécanismes inspirés du numérique facilitent l'accès au jeu comme culture et comme média.

Enfin, beaucoup des progrès réalisés par l'IA sont dus à la recherche en matière de jeu numérique (Dupuis, 2015). Or le joueur sur PC ou console n'a pas accès au code, alors que le joueur « de société » lit les instructions qui régissent son adversaire artificiel, il tire les cartes, il bouge les pions à la place de l'IA, en bref, il « soulève le capot ». En quoi ces dispositifs peuvent-ils contribuer à vulgariser les questions liées à l'IA ?

2. L'IA ANALOGIQUE À L'ÉPREUVE DES MODÈLES SÉMIO-COMMUNICATIONNELS

2.1 Le cadre théorique des sciences de l'information et de la communication

Pour chercher à répondre à nos interrogations, nous articulons ici nos observations avec l'analyse des mécanismes vidéoludiques qui, pour Genvo (2013), repose sur deux plans : celui des structures *du* jeu, qui sont liées au dispositif physique, à l'objet, et celui des structures *de* jeu, qui sont du côté de l'usage, de la pratique, de l'interprétation par le joueur.

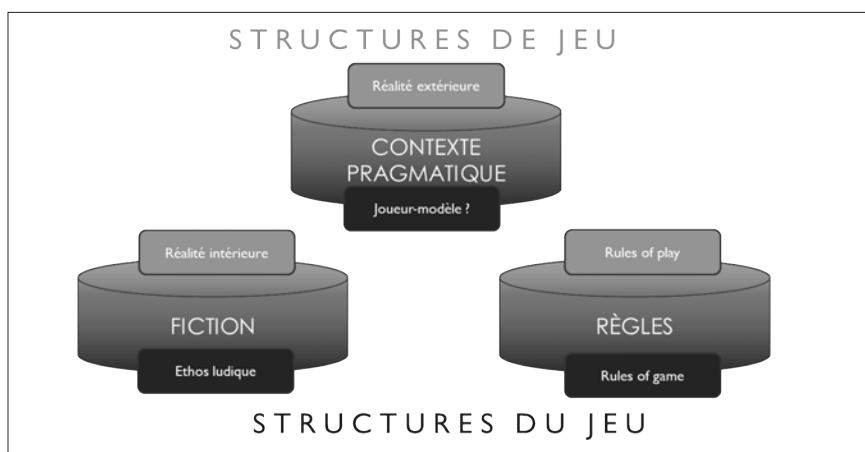
Ces deux plans coexistent au sein des trois instances déjà définies par Henriot (1989) : le monde fictionnel, les règles et le contexte pragmatique.

Le « monde fictionnel » s'incarne dans une « réalité intérieure » chez le joueur et repose sur « l'éthos objectif du jeu » (Genvo, 2018), révélateur de ce que le designer a à dire de l'univers fictif qu'il met en scène.

Le « contexte » repose lui aussi sur cette double incarnation : d'une part, une « réalité extérieure » vécue par le joueur et, d'autre part, le « joueur-modèle », c'est-à-dire le modèle d'usage que le concepteur a imaginé.

Enfin, des règles objectivement finalisées, les *rules of game*, s'articulent à leur actualisation par le joueur – les *rules of play* (voir figure 4.7).

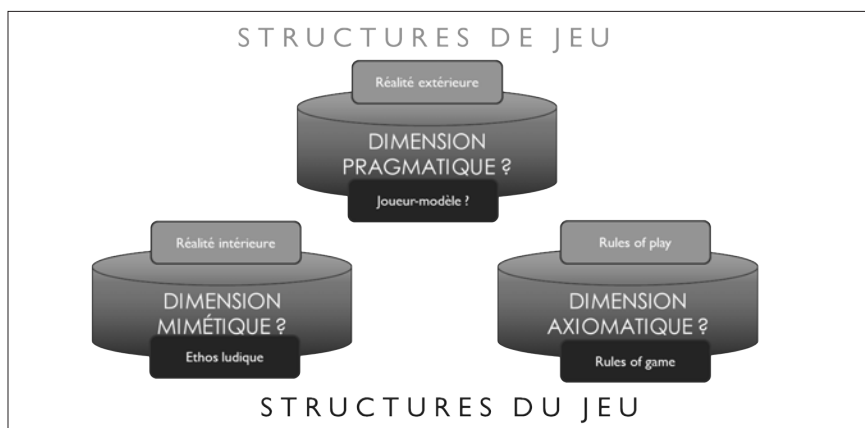
FIGURE 4.7
Monde fictionnel, règles et contexte pragmatique



Source : l'auteur d'après Genvo, 2013, 2018.

De son côté, Olivier Caïra (2016) considère que les dimensions du ludique mettent en coprésence une « esthétique mimétique », manière dont le dispositif simule la réalité, et une « esthétique pragmatique » : organisation concrète et ancrage dans la vie quotidienne. Les dimensions « mimétique » et « pragmatique » sont articulées par une « axiomatique », ou ensemble de règles et d'axiomes permettant de donner à l'expérience fictionnelle des implications concrètes pour le joueur, ou de représenter dans le monde fictif ses actes réels (voir figure 4.8).

FIGURE 4.8
Dimensions pragmatique, mimétique et axiomatique
et structure de/du jeu



Source : l'auteur d'après Caïra, 2016.

2.2 Un jeu, deux formes d'IA analogique

Mage Knight (2011) propose d'incarner un héros explorant un environnement qui se construit tuile après tuile. Le joueur affronte divers personnages grâce à l'utilisation habile de combinaisons de cartes (« combat », « influence », « repos »...), qui lui permettra d'en d'acquérir de meilleures au fil de la partie⁸. Pour impulser le rythme, la première édition du jeu intègre le mécanisme de « joueur fantôme ».

Dans le monde fictionnel de *Mage Knight*, le joueur fantôme n'a pas de « réalité intérieure ». Il n'est pas représenté sur le plateau et n'a aucune influence sur l'aventure vécue par le personnage du joueur. Un climat d'urgence est imposé par une pioche de cartes qui fait figure de chronomètreur (*timer*). Le « joueur-modèle » devra veiller à contrer le développement de cette pioche pour disposer de plus de temps pour réaliser ses actions. Absent du plateau, le « joueur fantôme » constitue un défi pour le joueur davantage que pour son personnage.

Quelques années plus tard, l'auteur du jeu propose un nouveau personnage permettant aux joueurs d'affronter un adversaire automatisé. Ce nouvel ennemi a une existence sur le plateau : il est représenté par une figurine. Il a

8. C'est la mécanique du *deckbuilding*.

un objectif dépendant du scénario : se rendre à un endroit, attaquer le joueur, etc. Ses actions sont régies par un mécanisme procédural à embranchements, fonction de la progression du personnage du joueur sur le plateau. Les « structures du jeu » qui l'animent sont moins prévisibles car elles dépendent des actions du joueur dans une « réalité intérieure » (voir figure 4.9).

FIGURE 4.9
L'ennemi automatisé de *Mage Knight*



Source : photographie de l'auteur.

S'il est possible à un joueur ayant joué contre cet adversaire de raconter sa partie, il sera probablement plus rare d'entendre un joueur narrer comment il a freiné la croissance de la pioche du joueur fantôme.

L'illusion d'un comportement *intelligent* du PNJ émerge lorsque le joueur peut donner un sens aux actions de l'ennemi artificiel. Il semble donc que « croire » en une IA revient à attribuer une signification à ses actions.

2.3 Un modèle sémio-communicationnel expérimental

En assimilant la dimension « fictionnelle » évoquée par Genvo à l'axe « mimétique » défini par Caïra, nous tentons ici de synthétiser ces deux topiques et une analyse systémique d'orientation sémiotique (voir Morgagni, 2011).

Dimensions axiomatique, pragmatique et sémantique, la proximité apparente entre le modèle sémiotique de Pierce (1935) et les topiques de

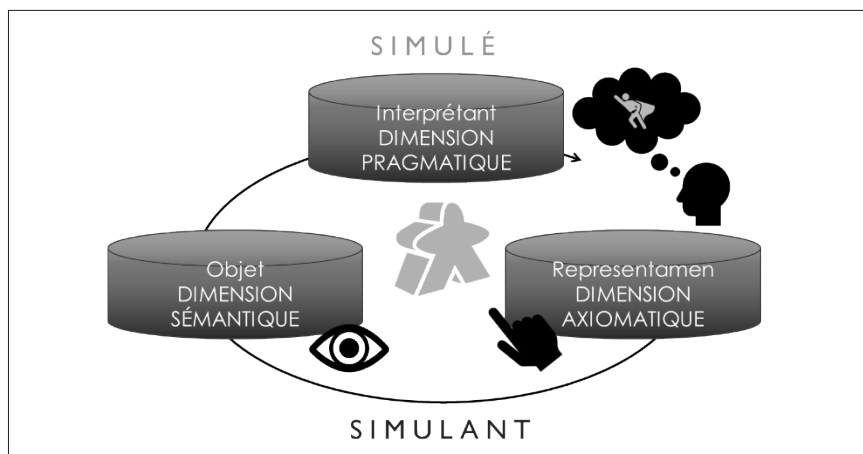
la recherche sur le jeu vidéo nous autorise maintenant à glisser de la *simulation* à la *signification*. Le simulat et le simulé deviennent ici les deux versants d'un processus de simulation, comme le signifiant et le signifié constituent les deux versants du processus de sémiose.

Pour C. S. Pierce, la signification prend forme de manière dynamique dans le processus de sémiose, actualisée par trois instances, dont nous avons cherché les traces dans les mécanismes ludiques de notre corpus.

Un mécanisme ludique est en effet une possibilité de signe. Cette possibilité s'exprime à travers (1) la manière dont est représentée une réalité extérieure au jeu ; (2) la façon dont cette représentation fait référence à l'objet ; et (3) la possibilité de son interprétation dans un contexte donné, en fonction d'un langage ou d'un système de logique donné.

Ainsi un *meeple*, petit personnage de bois symbolisant le joueur sur le plateau, désigne précisément son avatar ou sa faction⁹ ; il est signifiant – ou simulat – en ce qu'il renvoie à son objet par ressemblance physique¹⁰ ; il est interprété en vertu de l'expérience, de la logique qui a présidé à l'attribution des couleurs à chaque joueur¹¹ (voir figure 4.10).

FIGURE 4.10
Processus de sémiose et dimensions du simulat/du simulé



Source : l'auteur d'après Pierce, 1935.

9. Dans la topique piercéenne, on parlerait ici de « sinsigne ».

10. C'est, en termes piercéens, un objet de type « iconique ».

11. « Dicisigne » ou « signe dicent ».

Qu'est-ce que ce type d'approche peut nous aider à comprendre des mécanismes automatisés dans le jeu de société ? Nous nous pencherons, à titre expérimental, sur un second exemple de jeu à succès.

2.4 L'IA analogique au regard du processus de sémiase

Scythe (2017) met les joueurs à la tête de factions dans un univers retro-dystopique. Ils explorent le plateau, exploitent les ressources, construisent des bâtiments et déploient leurs unités.

À l'instar des joueurs humains, « l'automa » est à la tête d'une faction. Les mouvements et les stratégies de « l'automa » sont régis par une *carte-séquence*, organisant à chaque manche le tirage de *cartes-programmes*, qui à leur tour scriptent les actions des figurines de « l'automa » (voir figure 4.11).

La ligne 1 (« ligne d'action de déplacement ») demande au joueur humain d'activer « l'automa » en fonction de différentes conditions. Cette procédure se *décrypte* par un ensemble de ce qui constitue pour C.S. Pierce des « symboles » ; la simulation est rendue possible par des représentations traditionnelles et arbitraires ; son interprétation s'opère d'après les règles écrites, la structure du jeu. Le *choix de l'action* de « l'automa » peut être assimilé à une séquence de l'ordre de la « tiercéité¹² » (Pierce, 1935). Le joueur traduit cette ligne par : « si j'ai la puissance militaire suffisante, et l'occasion sur le terrain, j'attaque une unité importante du joueur humain ; sinon, et si j'en ai l'occasion sur le terrain, je conquiers un territoire occupé par un de ses ouvriers ; sinon je me contente d'avancer mes ouvriers ».

Dans le cas de la ligne 2 (gains de l'automate), l'objet (la pièce) fonctionne par ressemblance avec les pièces de carton. La ligne est interprétée par le joueur « directement » : l'IA récupère simplement deux pièces, sans que cela renvoie à une prise de décision de l'automa. La séquence est de l'ordre de la « priméité¹³ » et constitue une forme potentiellement moins *convaincante* pour les joueurs, au risque de représenter un programme de contraintes plutôt qu'un comportement signifiant.

Il semblerait que la crédibilité d'une IA soit liée à des processus sémiotiques d'ordres élevés, c'est-à-dire faisant appel à des « objets » désignés par convention plutôt que par ressemblance, à un « interprétant » lié davantage

12. Plus précisément, elle pourrait constituer un « légisigne symbolique argumental ».

13. On parlerait ici de « qualisigne iconique rhématique ».

au raisonnement qu'à la perception directe. En soumettant le mécanisme au risque de l'arbitraire, l'illusion d'intelligence croît à mesure que se complexifie le processus.

FIGURE 4.11

Plateau et figurines du jeu *Scythe* (en haut), cartes permettant de jouer les coups de « l'automata » (en bas)



Source : Éditions Stonemaier Games, 2017.

La seconde version de l'IA de *Mage Knight* et « l'automata » de *Scythe* sont conçus comme des *automates finis hiérarchiques* : ils possèdent plusieurs ensembles d'états. La transition d'un ensemble à l'autre s'opère en fonction d'une transformation de l'environnement. Le « joueur fantôme » de *Mage Knight* propose, lui, un défi aux joueurs ; il n'a ni but ni, bien sûr, de personnalité.

3. PERSPECTIVES POUR LA RECHERCHE-CRÉATION

Les mécanismes pour jouer avec un adversaire fictif dans le jeu de plateau s'apparentent donc à des règles de comportement (plus ou moins différenciées suivant les situations), mais ne proposent pas de fonction d'évaluation ou de mécanisme de rétroaction qui permettraient à cette IA d'adapter son programme en fonction de la réussite ou non des programmes précédents.

Les « algorithmes de planification », par exemple, sont des programmes permettant la recherche de chemin dans un graphe, entre une situation initiale donnée et un objectif à atteindre. Mieux adaptés à la planification à long terme que les méthodes *ad hoc*, ils peuvent être utilisés pour trouver le meilleur chemin dans un espace des états, par exemple dans un arbre de comportements, en adaptant l'algorithme de recherche par rétropropagation. Nous cherchons dans cette partie à puiser dans la recherche-crédation vidéoludique des possibilités nouvelles pour les concepteurs contemporains de jeu de société.

3.1 Algorithmes de planification et apprentissage machine

L'apprentissage machine (*machine learning*) est souvent utilisé dans d'autres aspects du développement des jeux vidéo, comme la génération procédurale de contenus et des décors, permettant par là de proposer des univers cohérents, mais aussi surprenants, car non déterminés, « avec une rejouabilité infinie » (Dupuis, 2015). Chaque actualisation génère donc ses normes propres et chaque ensemble de séquence sa spécificité.

Des dispositifs, ou activités ludiques analogiques, ont souvent été utilisés pour illustrer le *machine learning* et reproduire le comportement de joueurs qui apprennent un jeu. Ainsi, le dispositif MENACE (Matchbox Educable Noughts and Crosses Engine), a été conceptualisé et prototypé en 1961 par le chercheur britannique en IA, Donald Michie. Il s'agit d'un ordinateur mécanique fait de 304 boîtes (par exemple : d'allumettes) qui permet à un manipulateur de jouer contre un adversaire artificiel à une partie de morpion (*tic-tac-toe*, ou *noughts and crosses*). Les informations contenues dans les boîtes retournent à chaque étape du jeu le mouvement « sélectionné par la machine » et un processus de renforcement, positif ou négatif, permet à cette machine d'affiner ses coups de parties en parties, en fonction de l'issue de celles-ci (voir figure 4.12).

FIGURE 4.12

Machine simulant un adversaire qui apprend à jouer au Morpion



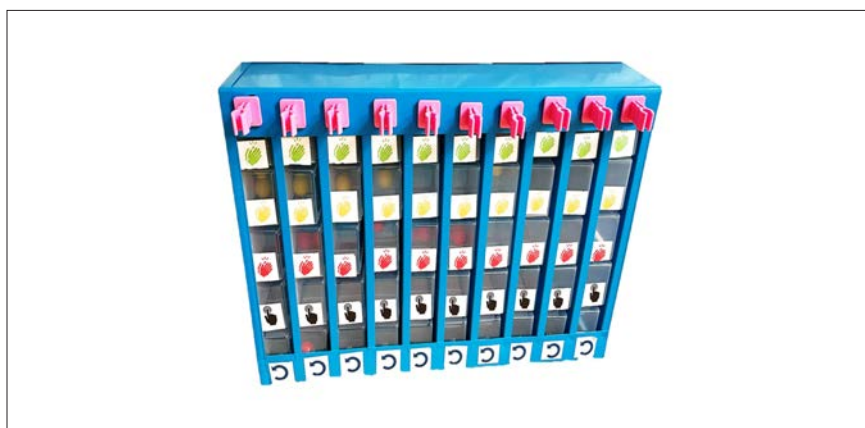
Source : réalisation et photographie de l'auteur.

Ainsi, *Hexapawn*, créé par un écrivain américain de vulgarisation scientifique, est une sorte de jeu d'échecs miniature sur un damier de neuf cases. Le faible nombre d'états (de parties possibles) permet de représenter simplement l'apprentissage machine au moyen d'un processus totalement mécanique (voir par exemple : Masse, 2019). Le dispositif consiste en un ensemble de boîtes où les possibilités de jeu sont représentées. Chaque boîte correspondant à une situation donnée sur le plateau, les *coups* de l'IA sont sélectionnés, parmi un ensemble de coups disponibles, en fonction de la situation présente. Les différentes possibilités de jeu sont retirées ou, au contraire, conservées dans la boîte correspondante en fonction de la victoire ou de la défaite de l'IA. Finalement, le joueur humain applique, lorsqu'il *fait jouer l'IA*, une procédure différente en fonction de l'échec ou de la réussite de ses coups passés. Le dispositif cherche donc, et sélectionne

par l'intermédiaire de l'être humain qui le manipule, un chemin toujours meilleur pour atteindre son objectif.

De manière similaire, différents dispositifs *apprenant* la pratique du jeu de Nim ont été imaginés par plusieurs groupes de recherche (voir par exemple : INRIA, 2011, 2012). Inspirés par le travail de Bell (2009) sur « l'informatique tangible », nous sommes actuellement en train de développer ce type de dispositifs analogiques en lien avec une ludothèque parisienne¹⁴ (figure 4.13).

FIGURE 4.13
Adversaire automatisé pour le jeu de Nim



Source : réalisation et photographie de l'auteur.

De leur côté, Collard, Henry et Hernalesteen (2020), ainsi que Briec, Henry, Smal et Frenay (2018) expérimentent avec des publics scolaires un jeu de plateau fonctionnant à la manière du « Qui est-ce ? ». La démarche consiste à faire concevoir aux élèves un jeu qui apprendrait *tout seul* à trouver les réponses aux questions des enfants, à partir des essais-erreurs de l'IA.

3.2 Utility-based AI

Avec les « approches basées sur l'utilité », plutôt que de passer d'un état à un autre en fonction des événements extérieurs, un agent va constamment évaluer les actions possibles et choisir l'action qui a la meilleure utilité présente. Pour chaque action est définie en amont une courbe d'utilité en

14. Voir : *L'Assoce*, <<http://www.lassoce.org>>.

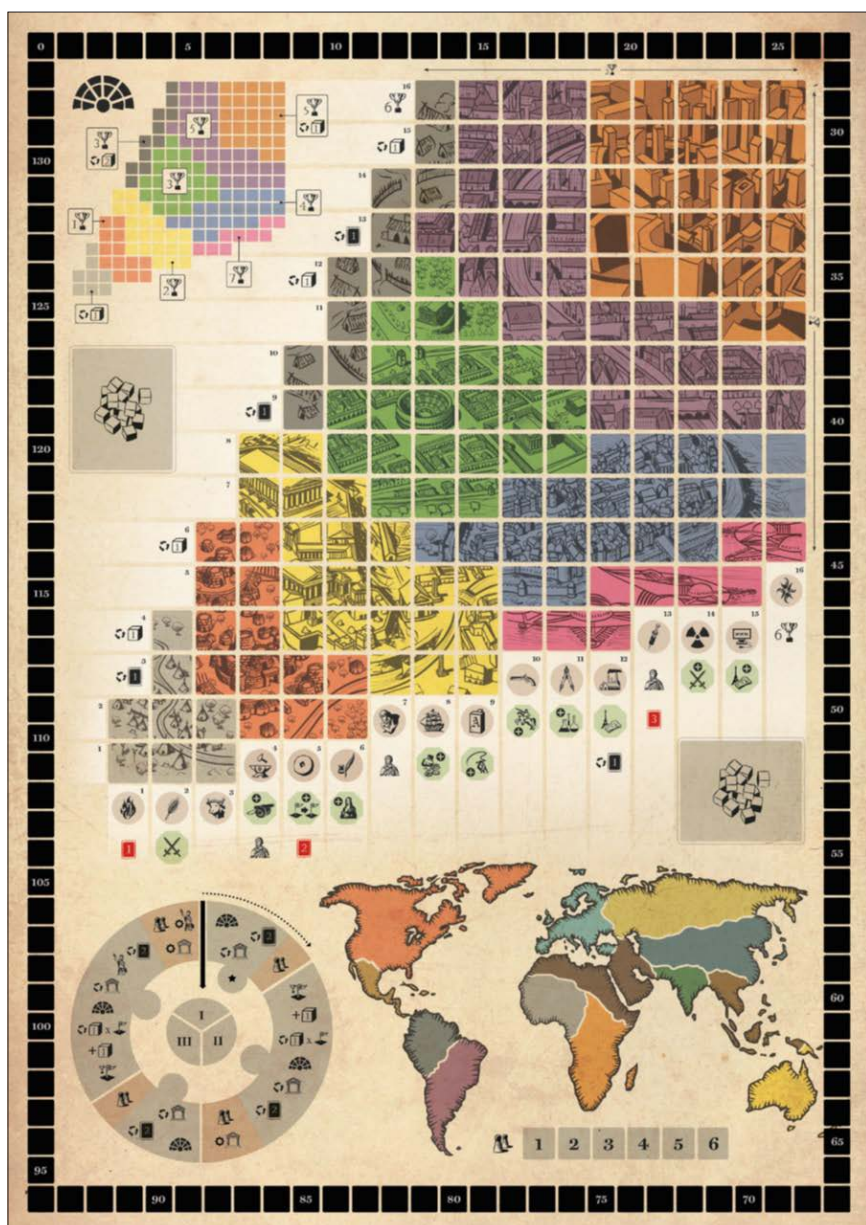
fonction de différents paramètres. Ces courbes déterminent la probabilité de choix de chaque action par le PNJ au regard de ces paramètres à un instant donné.

Dans *Historia*, les joueurs sont à la tête de civilisations et, tout comme les « civbots » (les civilisations automatisées qui permettent d'ajouter des joueurs virtuels), ils positionnent leur marqueur sur une matrice en fonction de leurs actions. En abscisse figure la progression scientifique de ladite civilisation, en ordonnée sa progression dans les « arts de la guerre ». L'emplacement du marqueur dans cette matrice indique le « type de gouvernement » adopté par cette civilisation : située dans la zone jaune, la civilisation est « féodale » ; dans la zone bleue, elle est « capitaliste » ; dans la zone rose, elle est « utopique », etc. (voir figure 4.14).

Les joueurs humains choisissent leurs actions à chaque tour au moyen de cartes, faisant éventuellement progresser leur civilisation sur l'axe scientifique ou sur l'axe militaire. Les « civbots » jouent des cartes tirées au hasard, selon un processus sémiotique simple qui ne laisse pas de place pour l'arbitraire. Les schémas comportementaux des « civbots » sont toujours identiques. Dès lors, il est envisageable de réfléchir à une manière de donner des programmes différents (des jeux de cartes différents) aux « civbots » en fonction de leur type de gouvernement : une civilisation féodale va peut-être jouer de manière plus agressive qu'une civilisation démocratique...

Toujours en lien avec des designers et ludothécaires, nous travaillons actuellement sur des matrices de comportements qui pourraient constituer une adaptation des courbes d'utilité pour le jeu analogique. L'état actuel d'un PNJ serait alors déterminé par plusieurs paramètres, représentés sur une matrice, au sein de laquelle des aires différentes définiraient des programmes de comportement particulier. Lorsque tel ou tel paramètre serait modifié, l'état du PNJ pourrait passer d'une aire à une autre, d'un programme à un autre. Différentes matrices pourraient être utilisées pour représenter différentes formes d'adaptation des PNJ à ces paramètres.

FIGURE 4.14
Plateau du jeu *Historia*



Source : photographie de l'auteur.

3.3 Des middlewares ou outils analogiques intermédiaires

Dans le monde numérique, la nécessité de combiner différentes méthodes (IA *ad hoc* et apprentissage machine) a conduit les développeurs à construire des logiciels médiateurs (*middleware*) face à l'étendue des calculs nécessaires. Il s'agit de couches « d'IA générique » transposables à différents contextes. Ces logiciels médiateurs déterminent ce qui sera commun et ce qui sera propre à chaque jeu (Dupuis, 2015).

Cayatte (2021) porte un projet de recherche et développement destiné à accompagner la conception de jeux vidéo à même de transmettre des émotions spécifiques. Il s'agit de créer des logiciels médiateurs permettant une transmission efficace et reproductible de schémas de comportement.

De la même manière, Tayari-Meftah (2013) a travaillé sur l'annotation des états, de façon à définir un modèle pour la représentation des émotions qui permet l'échange entre différents systèmes et l'application à différents dispositifs. Elle propose une approche de la représentation et de la communication des émotions selon laquelle toute émotion peut être « perçue » – en entrée – par un système par décomposition de vecteurs, combinaison linéaire « secondaire » de vecteurs « émotions de base » dans une matrice multidimensionnelle. En retour, l'émotion peut alors être exprimée – en sortie – par le système au moyen d'un vecteur, recombinaison selon le processus inverse.

La transposition de ce type de modèle au comportement des PNJ analogiques semble un terrain fécond pour la recherche-crédation. Pour Bustros (2009), en effet : « Nous constatons que le domaine de l'intelligence artificielle sur les jeux de société est très avancé en ce qui concerne les jeux classiques, mais pas sur les jeux de plateau » : il fait référence aux IA développées pour battre le joueur humain dans le jeu d'échecs (par exemple : *Deep Blue*) ou le jeu de go (par exemple : *Alpha Go*). Nous le rejoignons lorsqu'il poursuit : « Il serait intéressant, dans un projet futur, de créer des techniques générales, des modèles de conception et une documentation sur les jeux de plateau. Il y aurait assurément beaucoup à dire et à faire sur le sujet. »

4. L'IA ANALOGIQUE : PORTÉE ARTISTIQUE, SOCIALE ET HEURISTIQUE

L'approche sémio-communicationnelle et l'emprunt de notions issues des recherches vidéoludiques nous permettent-ils finalement de tirer quelques enseignements sur l'utilisation de l'IA dans le « jeu de table » ?

Il s'agit dans cette dernière partie d'étudier la question sous trois angles différents : la pertinence de ces mécanismes en matière de création, d'innovation ludique, les répercussions socioculturelles de ces innovations et la portée heuristique de ces applications de l'IA : peuvent-elles aider à faire progresser la connaissance, à vulgariser les mécanismes dont elles s'inspirent ?

4.1 Quelles perspectives en matière de création ?

Pour Collet (2018), l'objectif de l'IA appliquée au jeu vidéo est de créer un système qui permet de « contourner » des problèmes de conception. Les méthodes *ad hoc* proposent en effet une expérience de jeu fluide, nécessitant une quantité limitée de calculs, qui induisent une « illusion d'intelligence », rendent crédible et cohérent l'univers du jeu. Il s'agit le plus souvent de méthodes manuelles qui « ne peuvent pas réellement être définies comme de l'IA classique, mais [qui sont] considérées comme telles par l'industrie du jeu vidéo depuis son commencement ». En effet les joueurs attendent des PNJ un comportement tout de même assez formaté, qui autorise une maîtrise progressive du jeu.

Les mécanismes d'automatisation des jeux de société restent relativement simples : en l'absence de rétroaction, la réponse procédurale dépend peu des actions du joueur. Pourtant, se dessinent de nouveaux modèles, dont le jeu « analogique » hérite aujourd'hui de son cousin numérique : des structures conçues comme des « environnements libres où tout est possible » (D'Afflon, 2012).

La narration procédurale, ou construction du contenu faisant appel à des algorithmes plus ou moins élaborés, constitue une sorte de révolution dans le jeu vidéo (voir le travail de Ryan, 2007, Szilas, 2014 ou Giner, 2018) et amène à concevoir l'univers diégétique comme un « agent » au même titre que les adversaires qui le peuplent. Il en va de même dans le jeu de société : dans *Horreur à Arkham*, la succession des événements déclenchés dépend d'une pioche de cartes qui se construit en fonction des actions des joueurs. De la même manière, les « rencontres » que les joueurs sont susceptibles de faire dans chaque « colonie » de *Fallout* s'adaptent aux conséquences de leurs choix.

Les perspectives de création semblent se multiplier lorsque les designers empruntent des concepts à la recherche-crédation vidéoludique. Mais c'est moins dans le recours à un *matériel de jeu* numérique que dans l'utilisation des *procédures numériques de conception* que peut être améliorée

«l'agentivité» (*agency*) de ces récits en tant que lieux crédibles d'interventions de l'utilisateur dans l'intrigue, d'affrontement avec les personnages qui la peuplent.

D'un autre côté, si la culture numérique apporte aux professionnels du jeu des outils opérationnels ou conceptuels pour créer de nouveaux dispositifs, ces outils semblent également à même de répondre différemment aux enjeux sociaux de la pratique ludique.

4.2 Quelle est la plus-value sociale de l'IA des jeux analogiques ?

Avec l'intensité de l'activité ludique augmente de manière linéaire la pratique du jeu *en solo* ou *entre amis*, au détriment par exemple du jeu *en famille*, et les aspects «diégétiques» sont davantage recherchés (Francblu, 2020). Le jeu semble devenir une activité recherchée comme fin en soi, tributaire de la rencontre de pairs aux mêmes centres d'intérêt. C'est dans ce cadre que se développent des mécanismes qui permettent de transformer le jeu analogique en expérience plus ou moins automatisée, immédiate et souvent coopérative.

Les groupes se retrouvent aujourd'hui en parallèle de leur vie professionnelle ou familiale, en s'appuyant sur des dispositifs, des *structures de jeu* qui facilitent la mise en place, l'apprentissage des règles, l'intégration de personnes nouvelles. L'activité ludique prend donc place dans des contextes et des rythmes de vie en réponse à la transformation des attentes ludiques autant que sociales : les *jeux à usage unique* font de la partie une expérience non reproductible, mais immédiatement jouable, quels que soient les protagonistes autour de la table ou le contexte de leur rencontre. Les mécanismes d'activation automatique d'adversaires virtuels permettent de vivre une «aventure narrative» sans qu'aucun des participants n'ait à assumer le rôle chronophage de *maître de jeu*. La «sérialisation» fait vivre à des joueurs plus attirés par les aspects *épiques* et les objectifs à long terme des campagnes dont chaque épisode est l'occasion, le temps d'une soirée ou d'un week-end, de retrouver un groupe de pairs ou une communauté.

Or le jeu de société, comme tout produit culturel, peut être vecteur de représentations du réel en tant qu'il reflète des conceptions existantes, voire les étaye ou les fait évoluer¹⁵. En quoi ces dispositifs peuvent-ils faire

15. Nous avons par exemple contribué à un état des lieux des représentations des socio-écosystèmes dans le jeu de société, et des mécanismes ludiques à même de les faire évoluer (Francblu et Rotenberg, 2020).

avancer la compréhension de l'environnement *réel* du joueur ? Quel peut être le potentiel, en matière de médiation culturelle et de connaissance scientifique, de l'usage du jeu « analogique » ?

4.3 Quel est le potentiel heuristique de l'IA analogique ?

Dans les salles de classe ou dans les familles, différentes représentations et différents usages des outils numériques cohabitent (Bodart, Dumas et Henry, 2018). Les professionnels (ou les parents) restent peu armés pour démystifier les principes liés au fonctionnement des machines, les enjeux liés à l'IA, tandis que les jeunes s'emparent plus rapidement des outils technologiques. L'intégration du jeu vidéo dans l'offre éducative des ludothécaires ou des enseignants du primaire est encore en débat, en l'absence de balises claires ou de formations systématiques pour les professionnels.

Or il s'agit aujourd'hui de permettre aux enfants en particulier de se positionner davantage comme acteurs que comme consommateurs (Collard, Henry et Hernalesteen, 2020 ; Briec, Smal et Frenay, 2018). Car si la didactique du numérique est encore en construction, les œuvres de fiction autour de la cybernétique (Weiner, 1954) ou aujourd'hui les imaginaires qui touchent au numérique (Desbois, 2015) peuvent « empêcher le public de penser¹⁶ ».

Pourtant le jeu sérieux (Michael et Chen, 2006) garde l'empreinte de l'institution qui le met en place (Trépanier-Jobin, 2016). Il s'agit donc de proposer des expériences plutôt que de convaincre l'utilisateur. Le développement rapide de ce type d'approche dans un circuit du jeu vidéo indépendant amène Genvo (2016) à identifier des *expressive games*. Sans intention de persuader ou de convaincre, l'objectif est de partager des représentations que l'utilisateur peut s'approprier. L'activité ludique peut alors être utilisée à des fins praxéologiques en ce qui concerne la médiation scientifique, d'autant mieux lorsqu'elle est dégagée d'une ambition didactique trop évidente. Par exemple, Oddendal et Zavala (2018) ont conçu un jeu de plateau permettant *par ailleurs* de comprendre la structure algorithmique des réseaux sociaux.

Le jeu de société requiert, inévitablement, un ou des êtres humains pour mettre en place les règles, pour leur donner du sens et pour prendre

16. Pour Weiner, déjà en 1950, ni « l'insouciance optimiste » ni la « déploration technophobe » n'amènent à « ouvrir le capot du fonctionnement des machines ».

en main, littéralement, les pions, cartes, figurines, y compris lorsqu'il s'agit de manipuler un personnage adverse dont le comportement est régi par une IA. Parce qu'il est par essence médiatisé, les médiateurs du jeu que sont les ludothécaires peuvent contribuer à la transmission d'une certaine littérature numérique.

Grâce à un financement conjoint du ministère de l'Éducation nationale et de plusieurs organismes de recherche, la conception d'IA analogiques destinées à des actions de sensibilisation du grand public est en cours d'expérimentation. Nos premiers résultats devraient paraître à la fin de 2023.

CONCLUSION : L'IA OU L’AFFIRMATION DU JEU COMME PRODUIT CULTUREL

Si la révolution numérique a accéléré le mouvement de convergence entre les usagers et les producteurs de mondes fictionnels, l'intégration de supports numériques dans le jeu de société (ce que l'on nomme le « jeu hybride ») ne semble pas foncièrement modifier son « agentivité » (voir Caïra, 2014). Pourtant, comme en littérature, ces outils peuvent être l'occasion de faire évoluer les méthodes de conception (Bornet *et al.*, 2014), à défaut de transformer les usages.

Aujourd'hui, le jeu est (souvent) construit comme une narration et (de plus en plus) vécu comme une expérience ; l'intelligence artificielle ou, plutôt, les procédures de construction de l'intrigue qui s'en inspirent répondent aux mutations de la pratique. Or, à l'heure où l'on parle de « sciences participatives » (voir par exemple Bourg et Whiteside, 2010), expérimenter des outils issus de la rencontre entre le média ludique et la culture scientifique et technique apparaît aussi comme l'occasion de faire progresser la connaissance. Parce qu'elle permet de « voir le moteur », d'observer, de manipuler et de commenter son propre fonctionnement, l'IA analogique semble propice à la sensibilisation des publics à une approche réflexive.

C'est à travers le « design numérique » (Mauger, 2018) des supports matériels, plus que dans la numérisation de ces supports, que peuvent émerger des modèles qui réagissent aux actions du joueur, qui apprennent et qui déploient une existence autonome.

Et lorsque le jeu n'est plus simplement une interface, il génère une identité propre : si je ne joue pas mes coups, l'IA de *Scythe* aura tout de même la victoire. Capable de décisions révélant les choix diégétiques de

l'auteur, le jeu gagne-t-il alors le statut d'œuvre, en tant que son récit peut exister en l'absence de tout joueur ?

RÉFÉRENCES

- Bard, Nolan *et al.*, « The Hanabi challenge : A new frontier for AI research », *Artificial Intelligence*, vol. 280, 2019.
- Bell, Tim, Ian Witten et Mike Fellows, *Computer Science unplugged : l'informatique sans ordinateur, programme d'activités d'éveil pour les élèves à partir de l'école primaire*, 2009.
- Berry, Vincent, « Le renouveau du jeu de société », *La Revue du livre pour enfant*, n° 300, 2018, p. 124-129.
- Bodart, Antoine, Bruno Dumas et Julie Henry, « Programmation tangible pour les enfants : analyse de l'existant, classification et opportunités », *30^e conférence francophone sur l'interaction homme-machine*, Brest, France, 2018.
- Bornet, Cyril, Daniel De Roulet et Frédéric Kaplan, « La simulation humaine : le roman-fleuve comme terrain d'expérimentation narrative », *Cahiers de narratologie*, n° 27, 2014.
- Bourg, Dominique et Kerry Whiteside, *Vers une démocratie écologique : le citoyen, le savant et le politique*, Paris, Seuil, 2010.
- Brieuc, Olivier, Henry *et al.*, « Enseigner l'intelligence artificielle à des jeunes adolescents à travers l'informatique tangible : activité autour de la synthèse des émotions », *Didapro n° 7 – DidaSTIC*, HEP Vaud, 7-9 février 2018.
- Brody, Aymeric, « Les joueurs amateurs de poker : une communauté de pratique ? », *RESET*, vol. 4, 2015.
- Brougère, Gilles, « Du jouet industriel au jouet rationalisé », dans R. Jaulin (dir.), *Jeux et jouets : essai d'ethnotechnologie*, Paris, Aubier, 1979.
- Brougère, Gilles, *Jouets et compagnie*, Paris, Stock, 2003.
- Bustros, Alexandre, *Un système d'intelligence artificielle pour le jeu de plateau Axies & Allies*, mémoire présenté comme exigence partielle de la maîtrise en informatique, Université du Québec à Montréal, Service des bibliothèques, 2009.
- Caïra, Olivier, « Jeux vidéo et jeux d'interaction en face-à-face : vers un modèle unifié d'écologie de l'intrigue », *Cahiers de narratologie*, n° 27, 2014.
- Caïra, Olivier, « Théorie de la fiction et esthétique des jeux », *Sciences du jeu*, n° 6, 2016.
- Cayatte, Rémy, « La narration procédurale au service de la transmission d'émotions : principes d'émergence organique d'imprévus narratifs et émotionnels », *Colloque Lusor in Fabula*, Université de Rouen-Normandie, 2019.
- Cayatte, Rémy, « Projet Goblinz Story : bilan », *Expressive game : plateforme d'analyse et de création de jeux vidéo*, 2021, <<http://www.expressivegame.com/wp-content/uploads/2021/07/Goblinz-Story-Rapport-Final.pdf>> (consulté le 10 août 2021).
- Chauvin, Simon, *Un modèle narratif pour les jeux vidéo émergents : interface homme-machine*, thèse de doctorat en informatique, Paris, Conservatoire national des arts et métiers, 2019.
- Chouffot, Rémy, « Intelligence artificielle. Il y a un joueur dans le jeu ! », *Plato, le magazine des jeux de société modernes*, n° 120, 2019, p. 38-40.

- Coavoux, Samuel, « La carrière des joueurs de *World of Warcraft* », dans S. Craipeau, S. Genvo et B. Simonnot (dir.), *Les jeux vidéo au croisement du social, de l'art et de la culture*, Nancy, Presses universitaires de Nancy, 2010, p. 43-58.
- Collard, Anne-Sophie, Alyson Hernalesteen et Julie Henry, « Déconstruire les représentations médiatiques sur l'intelligence artificielle en jouant à *Qui est-ce ?* », *Actes du colloque Proceedings TICEMED 12*, Athènes, 7-9 avril 2020.
- Collet, Sébastien, « L'Intelligence artificielle dans les jeux vidéo », *Sagie*, 2018, <<https://www.saagie.com/fr/blog/blog-l-intelligence-artificielle-dans-les-jeux-video/>> (consulté le 1^{er} mai 2021).
- D'affon, Aymeric, « Sur l'échelle de la ludicité. Création et gamification », *Hermès, La Revue*, vol. 1, n° 62, 2012, p. 41-47.
- Dauphragne, Antoine, « La culture ludique : du jeu d'enfant au loisir adulte », *Actes du colloque Enfance et cultures : regards des sciences humaines et sociales*, 2010, <<http://www.enfanceetcultures.culture.gouv.fr/actes/dauphragne.pdf>> (consulté le 11 août 2018).
- Dauphragne, Antoine, « Le sens de la fiction ludique : jeu, récit et effet de monde », *Strenæ*, n° 2, 2011.
- Deledalle, Gérard, « Traduire Charles S. Pierce. Le signe, son concept et son usage », *TTR*, n° 3 (1), 1990, p. 15-29.
- Desbois, Henri, « A kind of magic, le rôle du surnaturel dans les fictions du numérique », *Interfaces numériques*, n° 4 (2), 2015, p. 223-236.
- Dupuis, Henri, « Intelligence artificielle et jeux vidéo », *Le site de vulgarisation scientifique de l'Université de Liège*, 2015, <https://www.reflexions.uliege.be/cms/c_384841/fr/intelligence-artificielle-et-jeux-videos> (consulté le 1^{er} mai 2021).
- Francblu, Samuel, *Pratiques du jeu de société contemporain*, janvier 2020, Paris, mis en ligne le 15 octobre 2021, <<http://www.lasocce.fr/fs/Root/eph11-Les-joueurs-GIS-L-ASSOCE.pdf>> (consulté le 1^{er} novembre 2021).
- Francblu, Samuel, « Pratiques du jeu de société : de l'expérience du play à la mutation du game », *Séminaire Temps, fictions et pratiques ludiques*, Besançon, France, 2020, <<https://mf.hypotheses.org/1422>> (consulté le 1^{er} novembre 2021).
- Francblu, Samuel et Miguel Rotenberg (2020), « Socio-écosystèmes et jeux de société. État des lieux », *Réseau des zones ateliers du CNRS-INEE*, 2020, <<https://www.lasocce.org/recherche-d%C3%A9veloppement>> (consulté le 1^{er} juillet 2021).
- Francblu, Samuel, « Les soirées jeux en ludothèque », *Plato, le magazine des jeux de société modernes*, n° 139, octobre 2021, p. 44-45.
- Francblu, Samuel, « Il était une fois. Les enjeux narratifs du jeu de société moderne », *Plato, le magazine des jeux de société modernes*, n° 134, avril 2021, p. 40-41.
- Genvo, Sébastien, « Penser les phénomènes de ludicisation à partir de Jacques Henriot », *Sciences du jeu*, n° 1, 2013.
- Genvo, Sébastien, « Defining and designing expressive games: The case of keys of a gamespace », *Kinephanos*, 2016, p. 90-106.
- Genvo, Sébastien, « Penser et analyser l'ethos ludique comme processus de médiation », *Colloque Entre le jeu et le joueur : écarts et médiations*, Liège Game Lab/Université de Liège, octobre 2018.
- Giner, Esteban, « La narration à n-corps », *Chroniques vidéoludiques*, 2018.

- Juul, Jesper, *Half-Real : Video Games Between Real Rules and Fictional Worlds*, Cambridge, The MIT Press, 2011.
- Henriot, Jacques, *Sous couleur de jouer*, Paris, Jose Corti, 1989.
- Hofstadter, Douglas et Daniel Denett, *Vues de l'esprit : fantaisies et réflexions sur l'être et l'âme*, Paris, InterEditions, 1987.
- Inria, Nancy Bidouille, « Les algorithmes », *Sciences manuelles du numérique : déconstruire l'informatique pour mieux la comprendre*, 2011, 2012, <<https://pixees.fr/deconstruire-linformatique-pour-mieux-la-comprendre-2/>>.
- Leorke, Dale, « Solo board gaming : An analysis of player motivations », *Analog Game Studies*, 2018, <<https://analoggamestudies.org/tag/solo-gaming/>> (consulté le 1^{er} juillet 2021).
- Marti, Marc et Raphaël Baroni, « De l'interactivité du récit au récit interactif », *Cahiers de narratologie*, n° 27, 2014.
- Masse Bastien, « Une initiation concrète à l'intelligence artificielle. Formation au numérique des élèves », 2019, <https://pedagogie.ac-clermont.fr/numerique-educatif/wp-content/uploads/sites/6/2021/12/Initiation_apprentissage_par_renforcement.pdf> (consulté le 1^{er} septembre 2023).
- Mauger, Vincent, « Design narratif : considérations préalables à son étude et à l'analyse de compositions ludofictionnelles sous le modèle EST », *Sciences du jeu*, n° 9, 2018.
- Michael, David et Sande Chen, *Serious Games : Games that Educate, Train, and Inform*, Tampa, Thomson Course Technology, 2006.
- Morgagni, Simone, « Repenser la notion d'affordance dans ses dynamiques sémiotiques », *Intellectica*, n° 55, 2011, p. 241-267.
- Murray, Janet, *Hamlet on the Holodeck : The Future of Narrative in Cyberspace*, Cambridge, The MIT Press, 1997.
- Odendaal, Adrian et Karla Zavala, « Black boxes out of cardboard : Algorithmic literacy through critical board game design », *Analog Game Studies*, 2018, <<https://analoggamestudies.org/2018/12/black-boxes-out-of-cardboard-algorithmic-literacy-through-critical-board-game-design/#comments>> (consulté le 1^{er} mai 2021).
- Pierce, Charles, *Écrits sur le signe*, Paris, Éditions du Seuil, 1978 [1935].
- Philippette, Thibault, « La dimension sociocognitive de la jouabilité », *Interfaces numériques*, n° 4 (1), 2015, p. 23-42.
- Plutchik, Robert, *Emotion, a Psychoevolutionary Synthesis*, New York, 1980.
- Ryan, Marie-Laure, « Jeux narratifs, fictions ludiques », *Intermédialités*, n° 9, 2007, p. 15-34.
- Simonian, Stéphane, Jean-Jacques Quintin et Sébastien Urbanski, « La construction des collectifs dans l'apprentissage collaboratif à distance : l'affordance socioculturelle des objets numériques », *Les Sciences de l'éducation – Pour l'Ère nouvelle*, vol. 49, n° 1, 2016, p. 63-90.
- Szilas, Nicolas, « Où va l'intrigue ? Réflexions autour de quelques récits fortement participatifs », *Cahiers de narratologie*, n° 27, 2014.
- Tayari-Meftah, Imen, *Modélisation, détection et annotation des états émotionnels à l'aide d'un espace vectoriel multidimensionnel*, thèse de doctorat, Université Nice Sophia Antipolis, 2013.
- Trépanier-Jobin, Gabrielle, « Differentiating serious, persuasive, and expressive games », *Kinephanos*, numéro spécial : Exploring the Frontiers of Digital Gaming : Traditional Games, Expressive Games, Pervasive Games, avril 2016, p. 107-128.

Wake, Paul, «Token gesture: Towards a theory of immersion in analog games», *Analog Game Studies*, 2019, <<https://analoggamestudies.org/2019/09/token-gestures-towards-a-theory-of-immersion-in-analog-games/>> (consulté le 1^{er} mai 2021).

Weiner, Norbert, *Cybernétique et société: l'usage humain des êtres humains*, Paris, Seuil, 1954.

Zabban, Vinciane, Nicolas Piñeros et Timothé Roux, «Financement participatif et engagement des consommateurs.rice.s dans l'essor du secteur du jeu de société: un "âge d'or" tiré par la demande?», *Sciences du jeu*, n° 14, 2021.

Ludographie

Les références ludographiques sont indiquées ainsi:

Titre, Auteur(s), éditeur originel et année d'édition

Andor, Menzel, Éditions Kosmos, 2012.

Comanantes, Hawthorne, Éditions Plaid Hat Games, 2019.

Dead of Winter, a crossroads game, Gilmour & Vega, Plaid Hat Games, 2014.

Détective, Trzewiczek, Portal Games, 2018.

Etherfields, Oracz, Awaken Realms, 2020.

Exit, Brand & Brand, Éditions Kosmos, 2017.

Fallout, Hajek & Fischer, Fantasy Flight Games, 2017.

Folklore the Affliction, Donovan & Blain, Greenbrier Games, 2017.

Gloomhaven, Childres, Cephalofair Games, 2017.

Historia, Pranzo, Éditions Giochix.it, 2014.

Horreur à Arkham 3^e édition, Wilson, Launius & Valens, Fantasy Flight Games, 2018.

Le Dilemme du Roi, Hach & Silva, Éditions Horrible Guide, 2019.

L'Île interdite, Leacock, Éditions Gamewright, 2010.

Les Demeures de l'Épouvante, Konieczka, Fantasy Flight Games, 2011.

Lords of Hellas, Kwapinski, Awaken Realms, 2018.

Mage Knight, Chvatil, Éditions Intrafin, 2011.

Mice and Mystics, Hawthorne & Bistro, Éditions Plaid Hat Games, 2012.

Okko Chronicles: le cycle de l'eau, Condette & Hub, Giga Mech Games, The Red Joker, 2019.

Outlive, Olivier, Éditions La Boîte de Jeu, 2016.

Pandemic, Leacock, Z-Man Games, 2009.

Scythe, Steigmeier, Éditions Stonemaier, 2017.

Tainted Grail, Piskorski & Swierkot, Awaken Realms, 2020.

The 7th Continent, Roudy & Sautter, Éditions Serious Poulp, 2017.

TIME Stories, Rozoy, Éditions Space Cowboys, 2015.

Unlock!, Demaegd, Éditions Space Cowboys, 2017.

Voyages en Terre du Milieu, Hajek & Holdinghauss, Fantasy Flight Games, 2019.

Xia: Legends of a Drift System, Miller, Far Off Games, 2014.

Zombicide, Raoult, Lullien & Guitton, Éditions Cool Mini or Not, 2012.

INTELLIGENCE ARTIFICIELLE, PROCESSUS CRÉATIF ET GÉNÉRATION PROCÉDURALE DE JEUX VIDÉO PAR CONCEPTION AUTOMATISÉE

Laura Iseut Lafrance St-Martin¹ et Maude Bonenfant²

Le jeu et l'intelligence artificielle sont liés depuis leurs débuts dans les années 1950 et n'ont cessé d'évoluer en parallèle (Yannakakis et Togelius, 2018). Le jeu *Nim* (Jørgensen, 2009 ; Rougetet, 2018) et le jeu de dames (Jackson, 2000) servent en effet de modèles pour le développement des premiers programmes informatiques précurseurs de l'intelligence artificielle. Plusieurs algorithmes fondamentaux, dont le très connu Minimax³, ont aussi été élaborés à partir du concept du jeu d'échecs (Togelius, 2019). Par la suite, les algorithmes d'intelligence artificielle développés dans l'industrie du jeu vidéo sont principalement utilisés pour adapter la difficulté

-
1. Professeure à l'École des arts numériques, de l'animation et du design (NAD), Université du Québec à Chicoutimi (UQAC).
 2. Professeure titulaire au Département de communication sociale et publique de l'Université du Québec à Montréal (UQAM).
 3. L'algorithme Minimax a été formulé pour la première fois par Alan Turing dans le cadre d'une automatisation du jeu d'échecs. Cet algorithme évalue l'ensemble des coups possibles et sélectionne celui qui avantage le plus sa position (*max*). Afin de calculer efficacement ce coup, il doit prévoir un certain nombre de coups à l'avance et doit aussi évaluer les coups de l'adversaire (en tenant pour acquis que celui-ci sélectionne le coup qui l'avantage aussi le plus (*mini*). Dans le cas du jeu d'échecs, la quantité de données à prendre en compte est énorme. Il a donc fallu attendre plusieurs décennies afin que les ordinateurs soient assez puissants pour pouvoir performer efficacement cet algorithme. Malgré le raffinement de plusieurs techniques de calcul Minimax depuis sa première formulation par Turing, le principe reste aujourd'hui le même.

du jeu aux compétences des joueurs et des joueuses ou pour animer les personnages non joueurs (*non-playing characters*) avec des techniques de détection de chemin (*pathfinding*) et d'automates finis (*finite state machines*) (Yannakakis, 2012). Ils servent également à générer procéduralement des environnements de jeu, de la musique ou encore des dialogues (*procedural-content generation*). Par exemple, dès 1980, le jeu vidéo *Rogue* (A.I. Design) utilise la génération procédurale d'environnements, d'items et d'ennemis de manière à offrir une rejouabilité virtuellement infinie aux joueurs et aux joueuses, principe qui a été repris dans de nombreux jeux tels que les célèbres *Minecraft* (Mojang, 2011) et *No Man's Sky* (Hello Games, 2016).

Certains jeux utilisent aussi des techniques d'intelligence artificielle plus avancées, comme l'apprentissage machine. Par exemple, dans le jeu *Black and White* (Lionhead Studios, 2001), les joueurs et les joueuses doivent entraîner une créature géante à performer certains comportements en la punissant ou en la récompensant. Le jeu *Left 4 Dead* (Turtle Rock, 2008) constitue un autre exemple de l'implémentation de techniques d'IA. Dans ce jeu, un *AI Director* analyse les comportements des joueurs et des joueuses, adapte la difficulté des parties en fonction de leur progression et varie les lieux d'apparition des ennemis. L'implantation des diverses instances faisant appel à l'intelligence artificielle a pour objectif de bonifier l'expérience vidéoludique en la rendant plus crédible, immersive et diversifiée. Ces techniques d'intelligence artificielle déterminent l'affichage du contenu ludique à un niveau local, c'est-à-dire dans chaque partie propre à un joueur ou une joueuse, mais n'influencent pas la conception du jeu comme telle.

Or, au tournant des années 2010, l'industrie vidéoludique connaît un changement profond dû aux avancées technologiques liées aux mégadonnées. En effet, plusieurs sphères de la société ont été transformées par le développement et la normalisation de la collecte, du stockage et du traitement de données en quantité massive de même que l'élaboration d'algorithmes sophistiqués qui en permettent le traitement autonome. Dans les années 1980 et 1990, l'industrie du jeu vidéo exploitait déjà des données produites par les joueurs et les joueuses, mais de manière limitée ou même manuelle. Toutefois, les avancées techniques et les transformations des modèles économiques de jeu à partir des années 2000 ont permis une collecte et une utilisation de plus en plus importante des données comportementales des joueurs et des joueuses. En effet, les plateformes de vente en ligne (notamment Steam), les plateformes mobiles de jeu, les consoles connectées à Internet ainsi que l'essor des jeux multijoueurs en

ligne (*multiplayer online games*) ont participé à la création d'un contexte propice à l'exploitation des données massives des joueurs et des joueuses dans le cadre du développement et de la commercialisation de jeux. À partir des années 2010, avec la puissance grandissante des ordinateurs, la baisse des coûts des serveurs et la vitesse augmentée de transmission des données, cette logique a pris une importance grandissante dans l'industrie. Aujourd'hui, la quasi-totalité des studios de développement de jeux vidéo utilisent les données comportementales⁴ des joueurs et des joueuses dans leur processus de création ou de mise à jour (Togelius, 2019, p. 80).

Appelée *the quantitative shift in the digital game industry* (Whitson, 2013), cette transformation à partir des années 2010 a eu des effets sur les plans épistémologique, ontologique, politique et éthique de la conception de jeux vidéo, surtout depuis que la génération procédurale de contenu ne se fait plus uniquement en circuit fermé. Les données comportementales des joueurs et des joueuses sont maintenant utilisées par les entreprises, parfois en temps réel, afin d'aider à la prise de décision pour le développement du jeu ou pour générer une partie du contenu (Summerville *et al.*, 2018). Même si encore relativement peu de techniques avancées d'intelligence artificielle sont utilisées dans la génération procédurale de jeux vidéo complets, plusieurs chercheurs et chercheuses (Nelson et Mateas, 2007; Hendrikx *et al.*, 2013; Treanor *et al.*, 2015; Togelius, 2019) dédient néanmoins leurs recherches à l'avancement des techniques de conception automatisée de jeux vidéo. Malgré le caractère fondamentalement conservateur de l'industrie du jeu vidéo, il s'agirait selon ces chercheurs et chercheuses d'une des avenues les plus prometteuses à la fois pour les jeux vidéo et pour l'avancement de la recherche générale en intelligence artificielle.

Dans ce contexte, l'objectif de ce chapitre est de problématiser certains des enjeux liés à une utilisation grandissante de l'intelligence artificielle pour appuyer ou transformer la création dans l'industrie vidéoludique jusqu'à concevoir entièrement un jeu de manière automatisée. Ainsi, il ne s'agit pas de présenter les résultats d'une étude empirique, mais plutôt d'offrir un cadre conceptuel à une future étude de terrain en s'appuyant sur les écrits de chercheurs et de chercheuses dans le domaine de l'intelligence artificielle, en études du jeu, en communication et en philosophie. Il s'agira donc, dans un premier temps, de positionner la conception automatisée de

4. Dans le contexte informatique, une donnée comportementale est produite lors d'une interaction avec une interface numérique, c'est-à-dire qu'il s'agit d'une information dans une base de données qui renvoie à une action de la part d'un utilisateur ou d'une utilisatrice.

jeux vidéo au sein des courants contemporains de l'intelligence artificielle et de l'intelligence computationnelle. Par la suite, nous présenterons la génération procédurale de jeux vidéo et donnerons des définitions de la création et des processus créatifs à partir des écrits de certains chercheurs et chercheuses dans le domaine. Finalement, nous soulèverons de manière synthétique des enjeux ontologiques, politiques et éthiques liés à l'utilisation potentielle de l'IA pour la production de jeux vidéo complets afin d'ouvrir des pistes de réflexion.

1. CONTEXTE ENTOURANT LA CONCEPTION AUTOMATISÉE DE JEUX VIDÉO

Actuellement, l'industrie du jeu vidéo a intégré les avancées en intelligence artificielle et en traitement automatisé de données principalement pour établir des analytiques du jeu (*game analytics*) (El-Nasr *et al.*, 2013). En collectant et en traitant des quantités massives de données sur les comportements des joueurs et des joueuses, les entreprises sont en mesure d'avoir accès à une grande quantité d'informations sur leurs consommateurs et consommatrices, toutes les plateformes de jeux étant désormais connectées à Internet. Lié à des techniques de mesure, le traitement de ces données vise surtout les métriques d'affaires (*business intelligence*) et les métriques de jouabilité (*game metrics*). Ce dernier type de métrique a pour but de calculer les usages des joueurs et des joueuses par rapport aux possibilités offertes par le jeu et peut maintenant servir, lorsqu'elles sont combinées avec d'autres types de techniques, à la conception automatique de systèmes de jeu complets par génération procédurale.

Dans leur recension de l'utilisation de la génération procédurale dans le domaine des jeux vidéo, Hendrikx et ses collaborateurs (2013) définissent six champs d'application : les unités (*bits*), les espaces, les systèmes, les scénarios, la conception (*design*) et les contenus dérivés (bandes dessinées, images, etc.). Dans ce chapitre, il sera question du cinquième type, c'est-à-dire de la conception automatisée de jeu. Développé à partir de la fin des années 2000, ce mode de génération procédurale de jeux complets n'en est qu'à ses balbutiements. Pourtant, comme nous l'avons spécifié en introduction, cette avenue est non négligeable pour l'industrie du jeu vidéo, dans la continuité des pratiques en lien avec la génération d'environnements, de musique et de dialogues.

Les systèmes de génération automatisée de jeux peuvent « réaliser une conception automatisée et intelligente de jeux, en raisonnant à la fois sur le système de règles abstraites du jeu et sur la réalisation visuelle de ces

règles» (Nelson et Mateas, 2007, notre traduction). Ces systèmes s'inscrivent dans l'approche de l'intelligence computationnelle (CI), laquelle est habituellement définie comme une partie de l'intelligence artificielle basée sur les méthodes dites «douces» : la logique floue (*fuzzy logic*), les réseaux neuronaux et la computation évolutive (IEEE Computational Intelligence Society, 2019). Contrairement aux méthodes «dures» utilisées dans certains autres champs de l'intelligence artificielle, l'intelligence computationnelle se base sur une imitation des processus cognitifs humains, ce qui signifie qu'elle introduit des éléments comme l'imperfection et des changements aléatoires. Bien que cette approche puisse sembler contre-intuitive pour un champ d'études basé sur les calculs et la gestion de données, ces techniques ont permis la résolution de problèmes auparavant insolubles⁵.

Sans prétendre remplacer complètement le concepteur humain du jeu, les systèmes de génération automatisée de jeux visent tout de même à être «en haut de la hiérarchie abstraite jusqu'à la position de méta-auteur ou méta-autrice du processus génératif qui produit les jeux» (Nelson et Mateas, 2007, notre traduction). Dans ce contexte, l'humain reste concerné, mais avec un rôle différent puisqu'il devient le juge des réalisations de l'IA : il détermine les critères d'évaluation de la réussite de l'algorithme d'intelligence artificielle (afin que l'algorithme puisse calculer son niveau de réussite ou administrer les récompenses ou les punitions).

2. LA CONCEPTION AUTOMATISÉE DE JEUX VIDÉO À PARTIR D'ALGORITHMES D'INTELLIGENCE ARTIFICIELLE

Par cette nouvelle rencontre entre les données massives comportementales et les algorithmes d'intelligence artificielle, surtout ceux qui sont en apprentissage machine, la conception des jeux et les processus de création seront encore plus affectés puisque, conformément avec la logique des analytiques de jeu, les informations produites par les comportements des joueurs et des joueuses influencent, voire dictent déjà des choix créatifs.

Rather than developing a complete product and releasing a full game, social games are usually released early as small games that are then incrementally built as the game becomes popular. This incremental improvement to the game can be informed

5. L'introduction de techniques de calcul douces est d'ailleurs ce qui a permis de concevoir un ordinateur capable de jouer efficacement au *Go*. Ce jeu traditionnel asiatique comporte en effet beaucoup trop de variables pour utiliser les mêmes techniques que pour les échecs (voir la note sur l'algorithme Minimax plus haut).

by players' actions in the current version of the game. This data driven approach (Rabin, 2000) emphasizes the use of data to understand and improve the game, rather than merely use the intuition of the designer (Isbister et Schaffer, 2008; Pagulayan et al., 2003; Derosa, 2007). (El-Nasr et al., 2013, p. 178)

Dans le cas d'un système s'adaptant aux comportements du joueur ou de la joueuse, le jeu produit par l'intelligence artificielle se conçoit comme une itération entre des comportements des joueurs et des joueuses (l'*input*) et l'algorithme de conception. Le jeu devient donc le résultat de cette procédure itérative entre le joueur ou la joueuse et l'intelligence artificielle où les comportements dictent les choix de contenu qui guident à leur tour les comportements, en fournissant chaque fois plus de données aux algorithmes qui peuvent alors s'adapter.

Dans d'autres cas, le rôle de l'intelligence artificielle est encore plus fondamental. Au-delà de l'adaptation d'un système de jeu aux comportements des joueurs et des joueuses, certains systèmes de conception automatisée (*automated game design*) possèdent la capacité d'aller puiser dans diverses banques de données (son, image, type de contrôle, etc.) afin d'assembler eux-mêmes l'ensemble du jeu. Bien que les critères de sélection soient établis par la conception de l'algorithme, l'intelligence artificielle possède une marge de manœuvre pour sélectionner un élément plutôt qu'un autre au sein de la base de données et pour l'agencer d'une manière précise parmi toutes les possibilités. Cette apparente liberté de choix mène plusieurs chercheurs et chercheuses (Colton et Wiggins, 2012) à soutenir que ces systèmes sont créatifs. Ce raisonnement computationnaliste ou cognitiviste est basé sur le principe que la cognition est une forme de computation : or, le substrat du « programme computationnel » n'a pas à être nécessairement un cerveau humain. Le test de Turing vise alors à déterminer si un système cognitivise ou non – en l'occurrence ici, s'il crée ou non. Ce raisonnement est similaire à celui d'Alan Turing dans l'élaboration de son *Imitation Game*, c'est-à-dire que l'apparence détermine la valeur du résultat, peu importe la procédure. Si un algorithme produit un jeu jugé créatif, alors l'algorithme lui-même est créatif, peu importe s'il est issu de l'expression d'une créativité humaine ou de calculs mathématiques. Cette logique fait appel au concept d'émulation⁶, ce qui signifie que le résultat doit être similaire à un original, peu importe la manière dont il a été réalisé.

6. L'émulation est conçue en opposition à la simulation. Alors que l'émulation vise un résultat semblable, la simulation propose une procédure semblable.

Ainsi, selon l'argument de Turing, si un algorithme produit un jeu qui est jugé créatif, il sera lui-même compris comme créatif.

Les travaux de recherche sur l'utilisation de l'intelligence artificielle pour créer des jeux vidéo ou des segments de jeux sont souvent animés par le désir d'un jeu infini (*endless game*). Togelius (2019) explique qu'il s'agit du rêve d'un moteur de jeu pratiquement autonome qui construit le jeu au fur et à mesure de l'exploration de l'univers par le joueur ou la joueuse. En se basant sur les comportements des joueurs et des joueuses, ce moteur produirait une expérience sur mesure qui correspond à leurs intérêts et à leurs styles de jeu. Ainsi, si le joueur ou la joueuse interagit plus souvent avec un type de contenu en particulier, le moteur du jeu s'adapterait pour produire plus d'expériences similaires. Le jeu s'adapterait donc aux préférences du joueur ou de la joueuse en temps réel de manière à offrir une expérience toujours nouvelle, optimale et personnalisée. Ce fantasme d'un moteur de jeu complet, sans fin et autonome, anime plusieurs des recherches récentes en intelligence artificielle dans le domaine des jeux vidéo.

Néanmoins, il est peu probable que les avancées récentes transforment aussi fondamentalement la conception de jeux vidéo à court terme⁷. Présentement, les nouvelles technologies d'intelligence artificielle ne sont pas rapidement adoptées dans l'industrie vidéoludique (bien que, par le passé, elle ait été à l'avant-plan des avancées en IA). Puisque les employés et les employées travaillent généralement sous pression et sont régis par des échéanciers très serrés (IGDA, 2019), l'implémentation de nouvelles technologies représente un risque pour les studios de développement (Cook, 2020). Ces derniers attendent généralement que le niveau de confort (facilité d'implémentation au sein de l'entreprise) et de crédibilité des technologies (potentiel prouvé d'apporter des bénéfices à l'entreprise) soit suffisamment élevé pour justifier un tel risque (Tidd et Trehwella, 1997).

Pourtant, les technologies liées à la conception automatisée de jeux vidéo représentent une avenue potentielle pour l'avenir de l'industrie du jeu vidéo, car elles radicalisent simplement la logique algorithmique déjà présente et croissante dans l'industrie. D'un point de vue historique, cette

7. Il existe un jeu de société créé par un algorithme d'intelligence qui a été commercialisé : *Yavalath* (2007). Ce jeu a été conçu en utilisant un langage spécifique, Ludi, fait pour représenter plus facilement les règles et les mécaniques de jeu. Néanmoins, la production de jeux vidéo est beaucoup plus complexe que celle de jeux de société et les mêmes technologies ne peuvent pas être utilisées pour faire les deux.

relation d'interinfluence entre les jeux vidéo et l'intelligence artificielle est une continuation cohérente basée sur leurs affinités. Dans *Playing Smart* (2019), Togelius affirme d'ailleurs que le futur des jeux vidéo se trouve dans l'intelligence artificielle et vice-versa. Il précise que les jeux vidéo sont particulièrement pertinents pour la recherche en intelligence artificielle puisqu'ils proposent des environnements propices à l'expérimentation en faisant appel à de nombreuses capacités cognitives humaines.

3. DÉFINITION DU PROCESSUS CRÉATIF DANS LE DOMAINE DE L'INTELLIGENCE ARTIFICIELLE

Sur le plan épistémologique, la désignation des résultats de procédures algorithmiques comme étant « créatifs » soulève des questions concernant le sens donné au mot « création ». Étymologiquement, la création est le résultat de l'action de créer, c'est-à-dire de « faire, réaliser (quelque chose qui n'existait pas encore) » (*Le Robert*). L'idée centrale de cette définition est donc la nouveauté, soit faire exister quelque chose d'inédit jusqu'à ce jour. À l'origine, le concept de création est fortement lié à un sens religieux (Dieu qui mène le monde à l'existence), mais, désormais, son sens plus laïque et commun renvoie simplement à la création d'un auteur ou d'une autrice.

Selon Togelius (2019), dans le contexte de l'intelligence artificielle, la créativité est conçue comme une recherche de solutions à un problème encodé, comme un manque de contenu. Cette manière de considérer la créativité se nomme la génération procédurale de contenu basée sur la recherche (*search-based procedural content generation*). Pour mener à bien sa recherche, c'est-à-dire pour trouver une solution à son manque de contenu, le système d'intelligence artificielle se base sur un mode de représentation du contenu (graphique, sonore, etc.), une base de données contenant l'ensemble des éléments pouvant être représentés et une fonction mathématique de réussite, c'est-à-dire des critères permettant d'évaluer le niveau de succès de la génération de contenu. Le résultat de cette recherche de contenu et de la combinaison des éléments est qualifié de « créatif ». Pourtant, cette vision simple de la créativité est problématique, car elle se limite à poser une équivalence entre « création » et « nouveauté ».

Certains chercheurs et certaines chercheuses utilisent néanmoins le terme « création » pour décrire les systèmes de conception automatisée,

comme le *Game-O-Matic*⁸ (Treanor *et al.*, 2012) ou *ANGELINA*⁹ (Cook *et al.*, 2012). Or, ces jeux sont en fait une recombinaison d'éléments précédemment connus. Pourtant, Togelius soutient avec défi que, « si "automatiser la créativité" peut sembler être un oxymore pour certains, la créativité n'est en fait ni plus ni moins automatisable que d'autres capacités cognitives humaines » (2019, p. 81, notre traduction). Cette manière de considérer non seulement la créativité, mais également les capacités humaines, s'inscrit dans une certaine vision de l'intelligence artificielle, notamment soutenue par Marvin Minsky (1988), où toutes les facultés humaines, jusqu'aux émotions, peuvent être émulées informatiquement : il s'agit du *computationnisme* qui est aussi associé au *Good Old-Fashioned AI* ou IA symbolique. Dans un article précédent, nous avons développé un contre-argumentaire à cette position controversée afin de démontrer, à partir de la théorie des affects, que le rapport affectif au monde, nécessaire à la créativité, est bien une faculté humaine impossible à simuler ni même à émuler (Bonenfant *et al.*, 2019). Boden, dans son article « Creativity and artificial intelligence » (1998), précise d'ailleurs que la créativité « implique non seulement une dimension cognitive (la génération de nouvelles idées), mais aussi la motivation et l'émotion, et est étroitement liée au contexte culturel et aux facteurs de personnalité » (1998, p. 347, notre traduction).

De plus, si la définition de « création » semble simple et consensuelle, la valeur accordée à ce qui est « créatif » est beaucoup plus ambiguë et discutable puisque le critère de « nouveauté » n'est pas suffisant pour relever d'un acte créatif. Il faut minimalement ajouter un critère de « valeur » : « Une idée créative est une idée nouvelle, surprenante et valable (intéressante, utile, belle...) » (Boden, 1998, p. 347, notre traduction). Or, cette valeur est parfois difficile,

-
8. *Game-O-Matic* est un système de génération de jeux conçu à l'Université de Californie à Santa Cruz. Afin de produire un jeu, ce système a besoin d'une liste d'objets ou d'acteurs ainsi que leurs relations respectives (fournies par un humain). À partir de cette conceptualisation, il propose des jeux ayant une représentation minimale. Plus précisément, le *Game-O-Matic* est un système spécialement conçu pour représenter des idées sous une forme ludique. Ainsi, il fonctionne en se basant sur ce que ces concepteurs appellent la « micro-rhétorique », ce qu'ils définissent comme « *a representational segment of gameplay within a videogame that could be supported by a convincing meaning derivation* » (Treanor *et al.*, 2012, p. 1).
 9. *ANGELINA* est un ensemble de systèmes de génération de jeux conçu à l'Imperial College à Londres. Plusieurs itérations différentes de ce système existent : *ANGELINA*₁, *ANGELINA*₂, *ANGELINA*₃, etc. Ils se basent tous sur l'analyse d'articles de journaux afin de produire des jeux de plateformes dans lesquels les joueurs et les joueuses doivent trouver des éléments (personnages, lieux, événements, etc.) pour reconstituer la nouvelle. Le système de génération détermine le type de graphique et de musique en fonction de son analyse de la nouvelle et des éléments présents dans ses bases de données.

voire impossible à formuler en mots, donc à traduire en langage informatique. Par exemple, le caractère « amusant » du jeu n'est pas toujours explicable.

Par ailleurs, être créatif renvoie au fait de formuler des schèmes de pensée inédits et non pas seulement de nouveaux assemblages à partir de schèmes existants. Boden distingue trois types de créativité en incluant les nouvelles combinaisons comme premier type de créativité (*combinational*) ; le deuxième type (*exploratory*) fait référence à de petites modifications sur les schèmes (*tweak*) ; quant au troisième type, il renvoie à des transformations des schèmes (*transformational*). Les deuxième et troisième types sont interreliés et parfois difficiles à distinguer : nous situons la créativité dans ce plan et non dans le premier type, car ce sont les règles elles-mêmes qui sont modifiées ou créées.

Le style d'un artiste exprime bien cette idée où l'agencement des éléments de base n'est pas la part créative de son œuvre, mais bien la manière dont il ou elle les agencera – donc le schème unique qui sera créé à partir des éléments déjà présents (couleurs de peinture, notes de musique, mouvements du corps, etc.). En d'autres mots, la créativité intervient lorsque les règles sont modifiées, et non pas à l'apparition d'un résultat inédit consistant en un nouvel assemblage de données produit à partir de règles existantes (ce qui reviendrait à copier le style d'un artiste). Si l'intelligence artificielle est ainsi limitée par les règles ou des procédures qui la définissent, elle l'est également par les données fournies à partir desquelles elle pourra produire de nouveaux assemblages : le système procédural ne créera pas de nouvelles données en tant que telles, mais que des variations à partir de modèles existants. Même s'ils sont très performants pour interpoler et effectuer certaines tâches précises à partir de leur ensemble de données d'entraînement, les algorithmes demeurent peu efficaces pour extrapoler à l'extérieur. Cependant, la recherche en apprentissage machine travaille actuellement sur cette limite, ce qui mènera éventuellement à évaluer le caractère créatif du système procédural.

Néanmoins, les processus créatifs humains présentent des particularités par rapport à la procédure de production machinique de jeux (Bonenfant *et al.*, 2018). Selon sa définition, la procédure se présente comme une suite d'étapes afin de produire un effet spécifique. Elle est similaire à la définition de l'algorithme qui, au plus simple, se définit comme un ensemble de règles opératoires propres à un calcul (*Le Robert*). Avec l'intelligence artificielle, le champ d'application du calcul est évidemment plus large, mais la logique demeure la même : il s'agit de définir (généralement en amont)

les étapes nécessaires afin de produire ou de réaliser un résultat – dont les paramètres, *a minima*, sont connus, bien qu'il puisse éventuellement s'adapter par « apprentissage ».

De son côté, le processus a un sens moins spécifique, même s'il est parfois utilisé de manière erronée comme synonyme de procédure. Toutes les étapes nécessaires à l'atteinte d'un résultat ne sont pas déterminées à l'avance dans un processus, qui s'actualise plutôt dans un flux continu de micro-événements qui l'affectent. Le flou, la variabilité et le dynamisme inhérents au processus contribuent à le distinguer de la rigidité, l'objectivation et la permanence de la procédure algorithmique. Bogost (2012) résume bien la différence entre les deux : « le processus implique le flux, tandis que la procédure implique l'opération ; le processus concerne les événements, tandis que la procédure concerne la logique ; le processus concerne le devenir extérieur, tandis que la procédure concerne l'essence intérieure » (2012, p. 10, notre traduction). L'explicitation de la procédure rend prévisible l'énonciation des résultats possibles, du moins théoriquement (car le cerveau humain est vite dépassé par les capacités informatiques). Au contraire, les résultats du processus sont virtuellement infinis puisqu'aucune règle *a priori* ne doit être explicitement formulée, car elles peuvent l'être au fur et à mesure du devenir du jeu.

L'approche processuelle est ainsi la seule manière de créer de nouveaux schèmes de jeu grâce à son indétermination dans les formes d'actualisation qui sont virtuellement infinies. Chez Deleuze (1968), en s'inspirant de Bergson, le virtuel se définit en opposition avec l'actuel, c'est-à-dire ce qui n'a pas (encore) pris forme dans le monde, qui n'a jamais été actualisé. Ainsi, le virtuel est ce qui ouvre sur les potentialités du monde et de son devenir. L'ordre du virtuel est la voie par laquelle de la différence est introduite dans le monde ; en opposition, l'ordre du possible fonctionne sur la base de répétitions où des occurrences ont eu lieu, théoriquement ou réellement. En ce sens, la part virtuelle des processus se distingue des procédures répétitives qui sont plutôt du domaine du possible. La conclusion rapide serait de lier les processus à la créativité humaine et les procédures, aux calculs machiniques.

Or, comme nous l'avons mentionné précédemment, certains courants de l'intelligence artificielle basée sur des méthodes dites « douces » relèvent d'une logique moins procédurale en adoptant une approche plus similaire aux facultés humaines. Bien que ces techniques d'IA ne soient que des métaphores biologiques qui ne reproduisent pas le réel fonctionnement de

la cognition humaine, l'introduction de variations aléatoires et de possibilités de réajustement remet en question notre compréhension du fonctionnement des algorithmes. Par exemple, le champ de l'intelligence artificielle nommée créativité computationnelle propose de concevoir d'emblée ces systèmes comme créatifs : le travail porte sur les règles et pas seulement sur la puissance d'agencement. Ce type de créativité machinique est défini comme « la philosophie, la science et l'ingénierie des systèmes informatiques qui, en assumant des responsabilités particulières, présentent des comportements que des observateurs impartiaux jugeraient créatifs » (Colton et Wiggins, 2012, notre traduction). Deux éléments clés sont au cœur de cette définition : la responsabilité et l'apparence de créativité.

D'abord, le système d'intelligence artificielle doit poser des actions dont la responsabilité lui revient et non pas à l'humain qui l'a conçu. Cette conception sous-entend donc une marge d'action et des choix pour la machine. En effet, la responsabilité est impossible dans un contexte entièrement prédéterminé qui n'offre aucun espace de libre arbitre – s'il est possible de parler de « libre arbitre » pour un système informatique. D'ailleurs, ce genre de raisonnement a justifié l'appel à l'interdiction des systèmes autonomes d'armes létales par près de 120 chercheurs et chercheuses en IA et en robotique en 2017¹⁰.

Ensuite, la créativité en elle-même n'est pas convoquée directement dans cette définition, mais plutôt sous la forme d'une *apparence* de créativité. Basés sur le même principe émulateur présenté précédemment, l'apparence et le résultat final priment sur la procédure (Colton et Wiggins, 2012). Ce type d'approche a mené au développement de nombreux systèmes qualifiés de créatifs par plusieurs, car ils travaillent à partir de formulations inédites de règles d'agencement de sons ou d'images¹¹. Si ces types d'algorithmes fonctionnent bien avec certaines formes de représentation, la complexité des jeux vidéo (combinant plusieurs types de contenus, comme des sons, des images, des règles mathématiques, etc.) fait en sorte qu'il est pour le moment impossible d'atteindre la même qualité en utilisant ce type d'outil d'intelligence artificielle.

Néanmoins, il semble pertinent d'analyser les effets que ces nouveaux champs d'application de l'intelligence artificielle auront sur notre

10. La lettre ouverte est disponible en ligne à l'adresse suivante : <<https://futureoflife.org/2017/08/20/autonomous-weapons-open-letter-2017/>> (dernière consultation le 31 janvier 2022).

11. Par exemple, le *Painting Fool* présenté par Colton et Wiggins a son propre « style ».

conception de l'acte créatif dans le domaine des jeux vidéo. Que ce soit au niveau des agents conversationnels, des comportements des personnages non joueurs, de la musique et des environnements générés procéduralement ou encore de l'adaptation du jeu au style des joueurs et des joueuses, les percées de la recherche en intelligence artificielle affectent l'industrie du jeu vidéo ainsi que l'expérience des joueurs et des joueuses. Même si l'intégration des techniques d'IA se fait relativement lentement, exception faite de l'analytique des jeux, la recherche constante d'innovations technologiques est un puissant moteur au sein du domaine vidéoludique. S'il n'est pas réaliste de voir, dans les prochaines années, des jeux commerciaux générés intégralement à partir de l'intelligence artificielle, il est fort probable que des assistants à la création soient intégrés dans les moteurs de jeu les plus utilisés (par exemple, Unity) (Togelius, 2019), rendant ces technologies de plus en plus communes.

4. LES ENJEUX DE L'UTILISATION DE L'INTELLIGENCE ARTIFICIELLE DANS LA CONCEPTION DE JEUX VIDÉO ENTIERS

4.1 Enjeux ontologiques

Si l'automatisation de la conception de jeux mène à contester la définition du processus créatif, elle remet également en question le statut même des jeux. Bien que le statut artistique des jeux vidéo soit aujourd'hui bien établi (Smuts, 2005 ; Tavinor, 2009 ; Kirkpatrick, 2011), l'ajout d'algorithmes intelligents ayant une autonomie décisionnelle dans la conception de jeux entiers complexifie de nouveau cette question. Nelson et Mateas (2007) croient que la conception automatisée de jeux déplace le statut d'œuvre du jeu vers le système de production lui-même. Ainsi, les jeux ne seraient plus considérés comme des œuvres, mais les systèmes les ayant conçus auraient acquis ce statut.

Si les procédures de conception acquièrent une valeur d'œuvre, le statut des ingénieurs et ingénieures et des programmeurs et programmeuses travaillant sur ces systèmes tendrait vers celui de créateurs ou créatrices, voire d'artistes. L'apparition et la normalisation potentielle des systèmes automatisés de conception viennent en effet modifier notre compréhension du statut artistique du créateur ou de la conceptrice : l'utilisateur ou l'utilisatrice de ces systèmes semble ne devenir qu'un exécutant avec une marge de liberté créative très réduite. D'ailleurs, certains et certaines dénoncent déjà la réduction de l'espace de création pour les concepteurs

et les conceptrices de jeux avec la présence de plus en plus grande de l'analytique des jeux qui dicte les choix créatifs sur la base d'une rationalité économique (Whitson, 2019).

Foucault, dans ses réflexions sur l'auteur (1969), propose une piste de réflexion afin de mieux comprendre le statut de créateur ou créatrice. Pour lui, l'auteur se distingue de l'individu qui écrit puisqu'il est plutôt une fonction dont bénéficient certains textes (qui peuvent être compris dans un sens large d'objets culturels). Certains textes, comme les contrats et les lettres, ont des signataires, mais n'ont pas d'auteurs à proprement parler. Ce genre de texte met en lumière le rapport problématique qui existe entre l'auteur et son œuvre. Toujours selon Foucault, l'auteur n'est pas la source de l'œuvre (contrairement à la vision romantique), mais plutôt un principe par lequel « on délimite, on exclut, on sélectionne » (1969). Cette fonction permet alors à la société de poser un jugement de valeur sur le texte et son utilisation. Audry et Ippolito appliquent d'ailleurs ce concept à une réflexion sur l'attribution de la valeur d'auteur dans les créations algorithmiques et soutiennent qu'« au lieu que la fonction d'artiste disparaisse lorsqu'une machine fait de l'art, elle peut être divisée en deux » (2019, p. 3, notre traduction). Le programmeur ou la programmeuse et l'algorithme acquièrent chacun et chacune une partie de la valeur d'auteur.

Cette conception foucauldienne de l'auteur appliquée au domaine de la production vidéoludique soulève la question des répercussions de l'utilisation grandissante de l'intelligence artificielle sur le statut social accordé aux concepteurs et aux conceptrices de jeux vidéo. Dans le contexte d'une production assistée par des algorithmes d'intelligence artificielle, l'humain conserve un important droit de regard en ce qui concerne la conception du jeu et l'acceptation des propositions algorithmiques (soit la formulation des critères de « valeur »). Or, un jeu conçu entièrement par un système d'intelligence artificielle brouille le jugement sur le responsable de la création. La fonction d'auteur dans un contexte d'intelligence artificielle demeure pour l'instant en suspens, mais elle aura des effets à la fois sur le jugement porté sur l'idée d'auteur et sur la responsabilité créative du jeu vidéo.

4.2 Enjeux politiques

Aujourd'hui, l'utilisation de l'IA dans la production vidéoludique est favorisée par des impératifs financiers venant de l'industrie. En effet, le contexte du développement de génération automatisée de jeux est, en grande partie, lié à une volonté de produire des jeux plus rapidement.

Par exemple, la naissance des *Newsgames*, forme journalistique utilisant les jeux vidéo comme support (Cook *et al.*, 2012 ; Cook et Colton, 2014 ; Treanor *et al.*, 2015), entraîne un besoin de rendre les jeux accessibles le plus vite possible après l'évènement à rapporter afin de conserver toute leur pertinence et leur pouvoir de sensibilisation sociale. Toutefois, de manière moins marginale, les producteurs et les productrices de jeux vidéo font généralement des pressions à l'accélération en automatisant une partie des tâches pour augmenter les profits en réduisant le coût (entre 20 et 150 millions par jeu) et le temps moyen de production des jeux (autour de cinq ans) (Hendrikx *et al.*, 2013, p. 2).

Ainsi, dans un contexte économique néolibéral d'efficacité et d'accélération, l'automatisation de la conception de jeux entraînera un changement sur le plan du travail et du statut des travailleurs et des travailleuses dans les studios de production vidéoludique (Commission de l'éthique en science et en technologie, 2019). Si, dans un futur proche, les algorithmes d'intelligence artificielle prennent plus de « responsabilités créatives », plusieurs questions sont soulevées par rapport à la reconnaissance du travail. Par exemple, l'agentivité créative des travailleurs et des travailleuses du milieu pourrait être évincée ou, plus réalistement à court terme, jumelée aux possibilités offertes par l'intelligence artificielle. Dans ce cas où plusieurs aspects de la conception et de la production de jeux vidéo seraient pris en charge par l'IA, demeure la question des métiers qui seront appelés à disparaître et des compétences nouvelles qui seraient requises par l'industrie – à commencer par une formation de haut niveau en informatique, alors que le milieu est déjà en pénurie de main-d'œuvre¹².

En fait, l'automatisation de certaines tâches, et plus encore de la conception complète du jeu, s'inscrit dans le paradigme de l'hypermodernité (Lipovetsky, 1983 ; Lipovetsky et Charles, 2004) où fleurissent de nouvelles formes de capitalisme telles que le ludocapitalisme (Dibbell, 2006). Les besoins du capitalisme, conjointement avec le néolibéralisme qui le nourrit depuis les années 1970, mènent à une accélération constante

12. Colin R. Singer, « Programmeurs, les entreprises canadiennes ont besoin de vous », *Immigration.ca*, 27 septembre 2018, <<https://www.immigration.ca/fr/computer-programmers-canadian-businesses-needs-you>> ; Gouvernement du Québec, « État d'équilibre du marché du travail à court et à moyen termes : diagnostics pour 500 professions », 2022, <<https://www.quebec.ca/emploi/informer-metier-profession/marche-travail/etat-equilibre>> ; « Pénurie d'informaticiens, et pas seulement en informatique », *Radio-Canada*, 23 novembre 2017, <<https://ici.radio-canada.ca/nouvelle/1067893/penurie-main-doeuvre-informatique-secteur-ti-programmeurs-developpeurs-numerique>>.

de la production, de la consommation et, conséquemment, de pans entiers de la société (Rosa, 2013) axés sur une jouissance individuelle toujours plus immédiate (Baudrillard, 1986). L'industrie du jeu vidéo s'inscrit dans cette logique où l'utilisation de l'intelligence artificielle est une des manières d'engranger plus de profits en plaisant au plus grand nombre possible de joueurs et de joueuses par la diversification et la personnalisation de l'offre à moindre coût. Ce contexte néolibéral doit être pris en compte lorsque les justifications de l'utilisation de l'IA sont évoquées, afin de ne pas viser l'augmentation des profits au nom d'une prétendue amélioration de l'expérience des joueurs et des joueuses. Cet argument est par exemple utilisé dans les jeux mobiles afin de collecter des données personnelles des joueurs et des joueuses (qui sont revendues sur le marché de la donnée) au nom d'une supposée « amélioration de l'expérience client ».

4.3 Enjeux éthiques

Outre les questionnements soulevés précédemment quant à l'utilisation des algorithmes d'IA pour produire des jeux complets, d'autres effets sur les joueurs et les joueuses, sur l'expérience de jeu et sur le sens véhiculé par les jeux sont à évaluer. Par leur statut de productions hybrides, créées par un système automatisé lui-même créé par des humains (pour l'instant), la valeur associée à l'expérience vidéoludique est modifiée puisqu'il n'est plus question de communication entre êtres humains (des concepteurs qui communiquent une certaine expérience vidéoludique à des joueurs et des joueuses), mais d'un rapport machinique entre le joueur ou la joueuse et le jeu. L'agentivité et la capacité d'appropriation, inhérentes à l'activité ludique, si elles ne sont pas plus limitées par le rapport procédural, sont certainement contraintes significativement puisqu'elles sont réduites à une communication humain-machine. Des études empiriques ont par exemple démontré que jouer contre des joueurs et des joueuses est plus satisfaisant (plus de plaisir, d'effet de présence et de sentiment d'expérience optimale [de *flow*]) que contre des avatars contrôlés par la machine (Weibel *et al.*, 2008).

Déjà, certains soulèvent les enjeux liés aux transformations de l'expérience vidéoludique alors qu'elle se rationalise de plus en plus à travers la médiation technologique et la standardisation des pratiques sociales à grande échelle dans des jeux vidéo populaires (Grimes et Feenberg, 2009). Dans ce type d'approche du jeu vidéo, l'individu a peu de marge de manœuvre pour négocier le sens et la valeur de son expérience puisque

presque tous les éléments sont définis à l'avance par l'objet technique ou par les joueurs les plus performants et les joueuses les plus performantes. À partir du moment où le jeu lui-même est conçu d'une manière seulement mathématique, ce rapport optimal, efficient, quantifié et rationnel apparaît en excès, de manière hyper-rationalisée¹³. Dans ce contexte, le processus de subjectivation des joueurs et des joueuses risque d'être aussi contraint, voire assujéti au sein d'une expérience dont le sens et la valeur sont prédéterminés par les paramètres de l'algorithme.

À terme, cette manière machinique de considérer le rapport au jeu vidéo mène à réduire les joueurs et les joueuses à des agents ou des agentes au sein d'une procédure, c'est-à-dire à des activateurs ou activatrices d'une suite d'actions à performer dans un certain ordre afin d'atteindre un état spécifique de « fin du jeu » avec le simple critère succès/échec (soit 0 ou 1). Ce type de conception cybernétique et procéduraliste de l'expérience vidéoludique a été critiqué par de nombreux chercheurs et de nombreuses chercheuses, particulièrement à partir des années 2010 (Arsenault, 2011 ; Adams et Dormans, 2012 ; Aarseth, 2014 ; Fabricatore, 2018). L'espace de liberté nécessaire pour s'approprier le jeu prévu par l'algorithme ne sera pas à envisager sur le même plan : poussée à son paroxysme, cette logique qui exclut la création de nouveaux schèmes de pensée rend les joueurs et les joueuses accessoires à l'activation du jeu puisqu'ils peuvent être eux-mêmes émulés.

De manière moins extrapolatoire et spéculative, une tension existe entre l'imprédictibilité du jeu produit par conception générative et la stabilité des règles nécessaire à l'exercice du jeu afin que les joueurs et les joueuses en comprennent le sens. En effet, certains jeux produits par intelligence artificielle sont activables, mais « injouables », et un observateur humain doit déterminer si le jeu est signifiant ou non. Par exemple, un algorithme pourrait créer un système mathématique incohérent, où les conditions de victoire seraient contradictoires avec les mécaniques de jeu, ou bien créer un « jeu » où aucun mouvement n'est possible pour le joueur ou la joueuse. Les assemblages des éléments sélectionnés par des algorithmes d'intelligence artificielle peuvent paraître insensés, hétéroclites

13. Même dans des jeux dont les règles sont très mathématiques, le facteur humain intervient dans l'actualisation du jeu. Un exemple célèbre est le Championnat du monde d'échecs de 1972 où l'impétueux et arrogant Bobby Fischer (États-Unis) a vaincu le discipliné et stoïque Boris Spassky (Russie). Les parties ne se sont pas entièrement jouées sur l'échiquier (Edmonds et Eidinow, 2005).

ou simplement non ludiques pour l'être humain qui interprète le jeu. Si le moteur de jeu infini présenté plus tôt est en théorie parfait puisqu'il produirait toujours une expérience ludique favorisant une immersion narrative plus profonde, il demeure pour l'instant un fantasme qui ne se réalisera pas dans un futur proche. D'ailleurs, il faudrait commencer par définir les paramètres de ce qu'est une « expérience ludique » afin de juger du « succès/échec » des propositions algorithmiques, ce qui reste matière à débat dans le champ des études du jeu.

À plus court terme est soulevée la question de l'expérience ludique produite par ses propres comportements puisque les techniques d'analyse de données comportementales sont déjà largement utilisées dans l'industrie vidéoludique (El-Nasr *et al.*, 2013). En collectant diverses données sur les actions en jeu, puis en les traitant en temps réel afin de modifier la réponse du jeu, ce système basé sur l'intelligence artificielle, entre autres par apprentissage machine, risque d'enfermer le joueur ou la joueuse dans ses propres modèles comportementaux, répétant ainsi au même les expériences passées. Cette manière de concevoir le jeu se comprend sous le même principe que la bulle de filtre (*filter bubble*) (Pariser, 2011), où la volonté des producteurs de plateformes de plaire le plus possible à leur public mène à une personnalisation du contenu conforme aux goûts et aux habitudes de chacun en les isolant au sein d'une consommation médiatique similaire et « réconfortante ». Dans ce contexte, la part virtuelle du jeu, perçue par le joueur ou la joueuse comme une expérience nouvelle, serait réduite afin de faire plaisir à chaque joueur ou joueuse, lui répétant ce que l'algorithme calcule de ce qu'il ou elle a précédemment aimé comme expériences de jeu.

Pour le moment, ce type d'approche où le jeu est modifié en fonction des actions des joueurs et des joueuses peut servir à détecter des modèles de jeu et changer la manière dont le jeu se déroule afin d'empêcher les joueurs et les joueuses de toujours reprendre les mêmes stratégies. Si cet usage bonifie l'expérience vidéoludique en la rendant plus stimulante, il demeure néanmoins une structure avec une logique potentiellement tautologique, cyclique et cybernétique où les comportements définissent l'expérience de jeu qui définit les comportements. Plus encore, l'hyper-personnalisation du jeu rendue possible grâce à l'apprentissage machine présente le danger d'un repli sur soi où l'autre sera exclu d'une expérience commune, qui peut être contrariante, mais aussi enrichissante. Après tout, jouer a toujours été une activité au cœur de la vie sociale (Huizinga, 1938; Caillois, 1958, etc.).

CONCLUSION

L'objectif de ce chapitre n'était pas d'analyser l'usage de plus en plus commun des algorithmes d'intelligence artificielle dans le domaine vidéoludique, mais plutôt de problématiser l'émergence d'une nouvelle manière de concevoir des jeux vidéo. Si, dès les années 1980, des jeux se basent sur la génération procédurale de certains contenus afin d'assurer une rejouabilité par l'introduction d'agencements nouveaux, les avancées technologiques de la dernière décennie ont néanmoins profondément modifié les pratiques de conception de jeux vidéo, notamment par l'utilisation grandissante des données massives comportementales pour guider les décisions de conception.

Plusieurs chercheurs et plusieurs chercheuses (Nelson et Mateas, 2007; Colton et Wiggins, 2012; Cook, 2017; Togelius, 2019) entrevoyent déjà une transformation des processus créatifs en y incorporant des algorithmes d'intelligence artificielle qui auront des « responsabilités » créatives, c'est-à-dire qui feront des choix sur le plan de la constitution des règles et des mécaniques de jeu ainsi que du rendu esthétique. D'un point de vue ontologique, l'intégration de plus en plus de techniques d'IA est susceptible de modifier notre compréhension à la fois de la création, de la valeur artistique des jeux vidéo et du statut de créateur des informaticiens et des informaticiennes. D'un point de vue politique, l'utilisation de l'IA s'inscrit dans une logique néolibérale d'accélération et d'augmentation des profits qui a des effets sur les travailleurs et les travailleuses de l'industrie, sur la reconnaissance de leur travail et sur le devenir des métiers. D'un point de vue éthique, les jeux conçus par des algorithmes d'IA peuvent aussi modifier l'expérience de jeu des joueurs et des joueuses, notamment en amplifiant le rapport hyper-rationnel aux règles. Une diminution de l'espace d'appropriation des joueurs et des joueuses au sein d'une communication exclusivement humain-machine risque de les enfermer dans leurs propres modèles comportementaux – manière confortable d'envisager les expériences vidéoludiques, mais également aliénante puisqu'elles sont isolées de toute vie sociale.

Cette réflexion quant aux rôles et responsabilités de l'intelligence artificielle est d'autant plus importante que les avancées dans le domaine des jeux vidéo auront nécessairement des répercussions sur d'autres domaines culturels et artistiques. Par exemple, plusieurs champs, comme le cinéma et la musique, entretiennent des rapports de proximité avec les jeux vidéo (Perron, 2008; Perron *et al.*, 2008; Blanchet, 2010;

Roux-Girard, 2011, 2014). De plus, l'industrie vidéoludique, forte de ses milliards en profit, investit massivement en recherche et développement informatique et sert de locomotive aux innovations, à l'acceptabilité et à l'adoption de nouvelles technologies. Considérant l'importance de ces productions culturelles dans la société contemporaine, un débat doit être ouvert afin de négocier les limites et de déterminer le rôle que nous voulons attribuer à l'intelligence artificielle et à ses « créations » dans notre vie culturelle. De nombreux jeux vidéo ont été créés en partie à l'aide d'algorithmes d'intelligence artificielle et ont réellement bonifié l'expérience des joueurs et des joueuses : l'idée n'est certainement pas d'écarter ces techniques procédurales dans la conception de jeux, mais de comprendre les effets de ces changements sur notre rapport processuel à l'expérience ludique et à la création.

RÉFÉRENCES

- Aarseth, Espen, « Ontology », dans M.J.P. Wolf et B. Perron (dir.), *The Routledge Companion to Video Games Studies*, New York, Routledge, 2014, p. 510-518.
- Adams, Ernest et Joris Dormans, *Game Mechanics : Advanced Game Design*, Berkeley, New Riders, 2012.
- Arsenault, Dominic, *Des typologies mécaniques à l'expérience esthétique : fonctions et mutations du genre dans le jeu vidéo*, thèse de doctorat, Université de Montréal, 2011.
- Audry, Sophian et Jon Ippolito, « Can artificial intelligence make art without artists? Ask the viewer », *Arts*, vol. 8, n° 1, 2019, p. 35.
- Baudrillard, John, *Société de consommation*, Paris, Gallimard, 1986.
- Blanchet, Alexis, *Des pixels à Hollywood. Pix'n Love*, 2010, <<https://univ-sorbonne-nouvelle.hal.science/hal-01370798>>.
- Boden, Margaret A., (1998), « Creativity and artificial intelligence », *Artificial intelligence*, vol. 103, n° 1-2, p. 347-356.
- Bogost, Ian, *Alien Phenomenology, or, What It's Like to Be a Thing* (s.l.), University of Minnesota Press, 2012.
- Bonenfant, Maude, Lucile Crémier et Laura Iseut Lafrance St-Martin, « Réflexions sémiotiques sur le circuit des données massives », dans André Mondoux et Marc Ménard (dir.), *Big Data et société : industrialisation des médiations symboliques*, Québec, Presses de l'Université du Québec, 2018.
- Bonenfant, Maude, Laura Iseut Lafrance St-Martin et Lucile Crémier, « Affected data : Understanding knowledge production in algorithmic events », *Global Media Journal*, vol. 11, n° 2, 2019, p. 66-78.
- Caillois, Roger, *Les jeux et les hommes*, Paris, Gallimard, 1958.
- Colton, Simon et Geraint A. Wiggins, « Computational creativity : The final frontier? », *Ecai*, vol. 12, 2012, p. 21-26.

- Commission de l'éthique en science et en technologie, *Les effets de l'intelligence artificielle sur le monde du travail*, 2019, <https://www.ethique.gouv.qc.ca/media/hqepc3fm/cest_effets_intelligence_artificielle_travail_a.pdf>.
- Cook, Micheal, « A vision for continuous automated game design », *arXiv:1707.09661 [cs]*, 2017, <<https://arxiv.org/abs/1707.09661>>.
- Cook, Michael, « Software engineering for automated game design », dans *2020 IEEE Conference on Games (CoG)*, 2020, p. 487-494.
- Cook, Michael et Simon Colton, « Ludus Ex Machina: Building a 3D game designer that competes alongside humans », *ICCC*, 2014, p. 54-62.
- Cook, Micheal, Simon Colton et Alison Pease, « Aesthetic considerations for automated platformer design », *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2012.
- Deleuze, Gilles, *Différence et répétition*, Paris, Presses universitaires France, 1968.
- Derosa, Phillip, « Tracking player feedback to improve game design », *Gamasutra*, August 7 2007, <<https://www.gamedeveloper.com/design/tracking-player-feedback-to-improve-game-design>>.
- Dibbell, Julian, *Play Money: Or, How I Quit My Day Job and Made Millions Trading Virtual Loot*, Basic Books, 2006.
- El-Nasr, Magy Seif, Drachen, Anders et Alessandro Canossa, *Game Analytics: Maximizing the Value of Player Data*, Springer, 2013.
- Edmonds, David et John Eidinow, *Bobby Fischer Goes to War: How a Lone American Star Defeated the Soviet Chess Machine*, Ecco, 2005.
- Fabricatore, Carlo, « Underneath and beyond mechanics. An activity-theoretical perspective on meaning-making in gameplay », dans Beat Suter, Mela Kocher, René Bauer (dir.), *Games and Rules: Game Mechanics for the « Magic Circle »*, New York, Columbia University Press, 2018, p. 87-111.
- Foucault, Michel, *Qu'est-ce qu'un auteur ?*, Paris, Gallimard, 1969.
- Grimes, Sara M. et Andrew Feenberg, « Rationalizing play: A critical theory of digital gaming », *The Information Society*, vol. 25, n° 2, 2009, p. 105-118.
- Hendrikx, Mark, Sebastiaan Meijer, Joeri Van Der Velden et Alexandru Iosup, « Procedural content generation for games: A survey », *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, n° 1, 2013 p. 1-22.
- Huizinga, Johan, *Homo Ludens : essai sur la fonction sociale du jeu*, Paris, Gallimard, 1938.
- Isbister, Katherine et Noah Schaffer, *Game Usability: Advancing the Player Experience*, Morgan Kaufman, 2008.
- IEEE Computational Intelligence Society, *What is Computational Intelligence ?*, 2019, <<https://cis.ieee.org/about/what-is-ci>>.
- IGDA, *Developer Satisfaction Survey*, 2019, <<https://igda.org/resources-archive/developer-satisfaction-survey-summary-report-2019/>>.
- Jackson, Michael, « Christopher Strachey: A personal recollection », *Higher-Order and Symbolic Computation*, vol. 13, n°s 1-2, 2000, p. 73-74.
- Jørgensen, Anker Helms, « An early Danish computer game » dans John Impagliazzo, Timo Järvi et Petri Paju (dir.), *History of Nordic Computing 2*, Berlin, Heidelberg, Springer, 2009, p. 283-288.

- Kirkpatrick, Graeme, *Aesthetic Theory and the Video Game*, vol. 1, Manchester University Press, 2011.
- Lipovetsky, Gilles, *L'ère du vide : essais sur l'individualisme contemporain*, Paris, Gallimard, 1983.
- Lipovetsky, Gilles et Sébastien Charles, *Les temps hypermodernes*, Paris, Bernard Grasset, 2004.
- Minsky, Marvin, *Society of Mind*, Simon and Schuster, 1988.
- Nelson, Mark J. et Michael Mateas, « Towards Automated Game Design », dans Roberto Basili et Maria Teresa Pazienza (dir.), *AI*IA 2007 : Artificial Intelligence and Human-Oriented Computing*, vol. 4733, 2007, p. 626-637.
- Pagulayan, Randy J., Kevin Keeker, Dennis Wixon, Ramon L. Romero et Thomas Fuller, « User-centered design in games » dans Julie A. Jacko et Andrew Sears (dir.), *The Human-computer Interaction Handbook : Fundamentals, Evolving Technologies and Emerging Applications*, Lawrence Erlbaum Associates, 2003, p. 883-906.
- Pariser, Eli, *The Filter Bubble : How the New Personalized Web is Changing What We Read and How We Think*, Westminster, Penguin Books, 2011.
- Perron, Bernard, « Genre profile : Interactive movies », dans *The Video Game Explosion : A History from Pong to PlayStation and Beyond*, 2008, p. 127-133.
- Perron, Bernard, Dominic Arsenault, Martin Picard et Carl Therrien, « Methodological questions in "interactive film studies" », *New Review of Film and Television Studies*, vol. 6, n° 3, 2008, p. 233-252.
- Rabin, Steve, « The magic of data-driven design » dans Mark DeLoura (dir.), *Game Programming Gems*, Charles River Media, 2000, p. 3-7.
- Rosa, Hartmut, *Social Acceleration : A New Theory of Modernity*, Columbia University Press, 2013.
- Rougetet, Lisa, « Machines designed to play nim games (1940-1970) : A possible (re)use in the modern French mathematics curriculum? », dans Eric W. Hart et James Sandefur (dir.), *Teaching and Learning Discrete Mathematics Worldwide : Curriculum and Research*, Cham, Springer International Publishing, 2018, p. 229-250.
- Roux-Girard, Guillaume, « Listening to fear : A study of sound in horror computer games », dans Mark Grimshaw (dir.), *Game Sound Technology and Player Interaction : Concepts and Developments*, IGI Global, 2011, p. 192-212.
- Roux-Girard, Guillaume, « Sound and the videoludic experience », dans Karen Collins, Bill Kapralos et Holly Tessler (dir.), *The Oxford Handbook of Interactive Audio*, Oxford University Press, 2014, p. 131-146.
- Smuts, Aaron, « Are video games art? », *Contemporary Aesthetics (Journal)*, vol. 3, n° 1, 2005.
- Summerville, Adam, Sam Snodgrass, Matthew Guzdial, Christopher Holmgård, Amy K. Hoover, Aaron Isaksen, Andy Nealen et Julian Togelius, « Procedural content generation via machine learning (PCGML) », *IEEE Transactions on Games*, vol. 10, n° 3, 2018, p. 257-270.
- Tavinor, Grant, *The Art of Videogames*, Chichester, John Wiley & Sons, 2009.
- Tidd, Joe et Martin J. Trehwella, « Organizational and technological antecedents for knowledge acquisition and learning », *R&D Management*, vol. 27, n° 4, 1997, p. 359-375.
- Togelius, Julian, *Playing Smart : On Games, Intelligence, and Artificial Intelligence*, Cambridge, MIT Press, 2019.

- Treanor, Mike, Bryan Blackford, Michael Mateas et Ian Bogost, « Game-O-Matic: Generating videogames that represent ideas », dans *Proceedings of the The Third Workshop on Procedural Content Generation in Games – PCG'12*, Raleigh, ACM Press, 2012, p. 1-8.
- Treanor, Mikes, Alexander Zook, Mirjam Eladhari, Julien Togelius, Gillian Smith, Michael Cook et al., « AI-based game design patterns », *10th International Conference on the Foundations of Digital Games*, 2015.
- Weibel, David, Bartholomäus Wissmath, Stephan Habegger, Yves Steiner et Rudolf Groner, « Playing online games against computer- vs. human-controlled opponents: Effects on presence, flow, and enjoyment », *Computers in Human Behavior*, vol. 24, n° 5, 2008, p. 2274-2291.
- Whitson, Jennifer, « The new spirit of capitalism in the game industry », *Television & New Media*, vol. 20, n° 8, 2019, p. 789-801.
- Whitson, Jennifer, « Gaming the quantified self », *Surveillance & Society*, vol. 11, n°s 1-2, 2013, p. 163-176.
- Yannakakis, Georgios N., « Game AI revisited », dans *Proceedings of the 9th conference on Computing Frontiers*, 2012, p. 285-292.
- Yannakakis, Georgios N. et Togelius, Julian, *Artificial Intelligence and Games*, New York, Springer, 2018.

PARTIE II

LE RECOURS À L'INTELLIGENCE ARTIFICIELLE ET AUX ALGORITHMES DE RECOMMANDATION POUR LA DIFFUSION DES CONTENUS JOURNALISTIQUES ET CULTURELS

6

LES OUTILS LIÉS À L'INTELLIGENCE ARTIFICIELLE EN JOURNALISME EN AMÉRIQUE DU NORD

Nicolas St-Germain¹ et Patrick White²

Le monde journalistique se trouve présentement à un carrefour. Deux changements de paradigme, le premier concernant la consommation médiatique des citoyens, et l'autre, l'économie désormais dite du numérique, affectent considérablement les médias traditionnels alors qu'ils doivent tenter d'innover sans cesse et de se faire une place dans l'échiquier numérique tout en vivant une crise financière permanente (Bizimana et Kane, 2019; Newman *et al.*, 2020). Si les médias doivent à ce point innover, c'est parce que le modèle d'affaires basé sur les revenus publicitaires est désuet. Plusieurs médias, généralement les plus riches, utilisent les nouveaux outils qu'offre le monde numérique, avec en tête l'intelligence artificielle (IA), pour remédier à la situation (Keefe *et al.*, 2021). C'est du moins ce qui est fait au quotidien canadien *The Globe and Mail* où l'outil Sophi gère 99 % des pages de son site Internet, en mettant de l'avant automatiquement et de manière personnalisée les contenus produits selon leur probabilité de rentabilité (The Globe and Mail Inc., 2020a, 2020b). Ultimement, l'outil a permis au média d'augmenter ses revenus d'abonnement et de réduire ainsi sa dépendance à la publicité (Turvill, 2021).

1. Gradué de la maîtrise en communication de l'Université du Québec à Montréal (UQAM).

2. Professeur agrégé de journalisme à l'École des médias de l'Université du Québec à Montréal (UQAM).

Si les initiatives utilisant l'intelligence artificielle reviennent à ce point dans les discussions quant à l'avenir des médias, c'est qu'historiquement la relation entre les innovations techniques et technologiques ainsi que le domaine du journalisme est une d'influence. Que ce soit la télévision, la radio, le téléphone ou même la photographie, le journalisme s'est développé avec l'apport de ces outils qui sont encore très présents dans le quotidien médiatique. Pour le chercheur et auteur John Pavlik, l'arrivée du téléphone n'a pas permis seulement de demeurer en contact malgré la distance, cela a « transformé la façon dont les journalistes rassemblent et rapportent l'information alors que plusieurs entrevues ont pu se faire par voie téléphonique³ » (Pavlik, 2000, p. 229). À ce sujet, l'historien Christian Delporte résume efficacement cette relation entre innovation et journalisme :

L'information s'approprie l'innovation technique, la détourne de sa destination première, l'applique à ses besoins ; car l'information, par ses caractères mêmes, doit impérativement se saisir des derniers progrès dans le domaine de la technologie de la communication (Delporte, 2005, p. 204).

Pour John Pavlik, chaque changement technologique cause quatre transformations dans le domaine (2000, p. 229). Premièrement, cela transforme la façon dont les journalistes effectuent leur travail, ce qui vient, dans un deuxième temps, affecter le produit médiatique en changeant autant son contenu que sa forme. Ces deux premières influencent dans un troisième temps la structure organisationnelle des salles de nouvelles. Quatrièmement, la relation entre les groupes de presse, les journalistes et le public se voit modifiée. Ces modifications se perçoivent également dans les salles de rédaction intégrant des outils liés à l'IA alors que cette nouvelle technologie apporte de nouvelles façons pour le journaliste de travailler – ce que l'on retrouve dans le journalisme algorithmique abordé par la chercheuse Laurence Dierickx (2021a) –, ce qui a des répercussions certaines dans le contenu couvert alors qu'il est possible aujourd'hui d'utiliser la technologie pour automatiser la récolte de données et générer des articles automatiquement. De plus, tout cela revoit la façon dont les salles de rédaction sont structurées alors que l'ajout d'expert en données et la constitution d'équipes multidisciplinaires en recherche et développement se font de plus en plus fréquents (The New York Times, s. d. ; Turvill, 2021). Finalement, la relation entre les médias et le public se transforme, étant

3. Traduction libre de « *it also transformed how journalists gather and report the news, frequently conducting interviews by telephone, and sometimes distributing news via the telephone* » tiré du texte de John Pavlik disponible en référence.

maintenant filtrée par des algorithmes, principalement ceux des GAFAM (Google, Amazon, Facebook, Apple, Microsoft), qui sont au cœur du monde numérique. De même, et ce chapitre en fera l'analyse, la relation entre les médias eux-mêmes se voit changée alors que de plus en plus d'initiatives de collaboration avec les universités et les entreprises en démarrage (*startups*) se mettent en branle pour partager les connaissances quant à l'avenir du domaine face aux changements causés par l'arrivée de ces outils (London School of Economics and Political Science, s. d. ; Sandalow, 2021).

Par ailleurs, ces changements vécus à l'intérieur des salles de rédaction, au fil des ans, ont aussi eu lieu à la suite de l'intégration d'Internet. En effet, cette dernière a eu son lot de répercussions sur le domaine alors que, depuis une vingtaine d'années, les organisations médiatiques ont connu trois phases transitoires selon le directeur du groupe de réflexion Polis à la London School of Economics, Charlie Beckett (2020). Tout d'abord, au tournant des années 2000, les entreprises ont transité vers le monde en ligne en se créant des sites Internet. Ensuite sont apparus les réseaux sociaux entre 2005 et 2012 où tous les médias ont dû se créer des pages. Et, dernièrement, le chercheur mentionne que désormais le journalisme entre dans une phase où les articles seront de plus en plus distribués, où l'on comptera plusieurs organisations de presse, et où les façons d'obtenir les nouvelles se feront de plus en plus nombreuses. C'est dans cette dernière phase que les outils liés à l'IA entrent en jeu. En effet, ceux-ci ont la capacité d'intervenir dans différentes étapes de la production journalistique et ainsi de faciliter le travail des journalistes.

Toutefois, comme nous l'avons mentionné, la situation financière des médias semble intimement liée à cette capacité pour les médias traditionnels d'intégrer ces outils. En effet, pratiquement toutes les initiatives présentées dans ce texte proviennent d'entreprises médiatiques dont la portée est substantielle. Le rapport *New Powers, New Responsibilities: A Global Survey of Journalism and Artificial Intelligence* constate une inégalité d'accès à la technologie (Beckett, 2019). Une réalité confirmée par une étude de la Knight Foundation (Keefe *et al.*, 2021), produite à partir de 130 projets intégrant des outils liés à l'IA où une disparité entre les médias nationaux et locaux est notée. C'est aussi ce qui ressort de notre étude intitulée *Les impacts de l'intelligence artificielle sur les pratiques journalistiques au Canada* : neuf des treize répondants mentionnent avoir besoin d'une aide financière pour concevoir des outils liés à l'IA (St-Germain et White, 2021). Combler cet écart qui semble se creuser est l'un des objectifs de la Knight Foundation en lançant une initiative de trois millions de dollars

pour permettre aux médias locaux d'utiliser des outils liés à l'IA (Cheung, 2021). Par ailleurs, pour pallier ce problème d'accessibilité, la collaboration entre les médias, les universités et entreprises en démarrage (*startups*) apparaît comme une piste de solution.

Face à tous ces changements que doivent affronter les médias, ce chapitre a l'intention d'aborder l'*intelligence artificielle* et plus précisément les outils associés à cette technologie qui, selon nos recherches, ont comme potentiel d'aider les journalistes et les entreprises médiatiques au quotidien en les libérant d'un travail routinier, dans plusieurs secteurs du domaine. Le but est de permettre aux journalistes de se concentrer sur des contenus à haute valeur ajoutée, tels que du journalisme de données ou de solutions, de grandes entrevues, de longs reportages, des balados (*podcasts*) ou tout autre contenu journalistique que les outils liés à l'IA ne peuvent accomplir. Dans le but de faciliter la compréhension de cet article, le texte se découpera en deux parties. Dans un premier temps, l'intelligence artificielle comme concept sera abordée pour bien montrer ce qui est attendu de cette technologie lors de son intégration. Dans un deuxième temps, l'initiative du Knight Foundation Lab sera présentée, de même que quelques projets mis sur pied par différents médias aux États-Unis. Tout au long de cette deuxième section, des liens seront faits avec notre recherche effectuée à l'automne 2020 au Canada où certaines craintes de nos répondants ont été formulées. Finalement, la conclusion reviendra sur les éléments centraux, soit le besoin de collaboration, le phénomène d'hybridation⁴, de déconstruction⁵ et l'embauche d'experts en informatique tout en abordant le concept des cinq V – volume, *velocity* (fréquence), variété, valeur et véracité – lorsqu'il est question de traiter et d'utiliser des données pour concevoir ces outils.

1. DÉFINITIONS DE L'IA ET SON LIEN AVEC LE JOURNALISME

L'intelligence artificielle comme secteur du domaine informatique ne possède aucune définition faisant l'unanimité. Pour certains, le terme « désigne des formes de vie artificielle qui peuvent surpasser l'intelligence humaine, tandis que, pour d'autres, toute technologie de traitement de données, ou presque, peut être qualifiée d'IA », est-il écrit dans le cours d'introduction *Elements of AI* (Université d'Helsinki et Reaktor, s. d.). En

4. *Hybridization* (Knight Center Courses, 2019a).

5. *Decomposition* (Knight Center Courses, 2019b).

somme, il s'agit plutôt d'un « terme un peu fourre-tout pour désigner les nombreuses possibilités offertes par les récents développements technologiques » (Dierickx, 2021b). En ce sens, les possibilités qu'offre cette technologie sont nombreuses. Elles incluent non exclusivement l'automatisation, soit un « système qui effectue une tâche précédemment accomplie de façon complète ou partielle par un humain » (Knight Center, s. d. ; traduction libre) et l'apprentissage machine – *machine learning*, en anglais – qui se veut un « algorithme permettant à l'ordinateur d'apprendre et de détecter des tendances dans les jeux de données » (Knight Center, s. d. ; traduction libre ; Beckett, 2019). Lorsque cet apprentissage est appliqué au journaliste, on inclut les outils liés à l'IA dans trois champs, soit la collecte et l'analyse d'information (1), la production (2) et la distribution (3). Par ailleurs, chaque salle de rédaction utilisant la technologie tend à définir l'IA selon l'utilisation qui en est faite (Beckett, 2019). À ce sujet, un document produit par le groupe de réflexion Polis de la LSE précise que cette pratique est même recommandée afin de mieux établir une stratégie organisationnelle pour l'usage des outils liés à l'IA (Beckett et Peretti, 2019, p. 11).

Pour cette raison, nous avons aussi statué sur une définition que nous croyons efficacement répondre aux attentes de la technologie dans le domaine journalistique. Par conséquent, lorsqu'il sera question d'intelligence artificielle, nous ferons référence à la définition donnée par le directeur du Computational Journalism Lab à l'Université Stanford, Nicholas Diakopoulos : « L'IA est un programme informatique en mesure d'effectuer des tâches qui nécessitent habituellement un certain niveau d'intelligence humaine⁶ » (Kelly, 2020). En ce sens, l'expression « outils liés à l'IA » fait référence à des programmes automatisés venant aider le journaliste dans ses tâches au quotidien de la même façon que l'ordinateur et la photographie l'ont fait lors de leur intégration. Ces deux définitions viennent illustrer ce qui est attendu de cette technologie, soit une assistance au travail journalistique et non pas un remplacement des effectifs. De même, nos lectures sur le thème de l'*intelligence artificielle* appliqué au journalisme semblent conclure que le terme *intelligence augmentée* répond mieux aux visées de la technologie, lorsqu'elle est utilisée dans les salles de rédaction. Comme le précise Francesco Marconi, l'auteur du livre *Newsmakers: Artificial Intelligence and The Future of Journalism*, et Alex Siegman : « Lorsque les journalistes utilisent des outils automatiques pour

6. Traduction libre de : « A.I. is a computer system that can perform a task that would typically require some level of human intelligence ».

bonifier leur reportage ou leur texte, nous appelons cela du journalisme augmenté⁷ » (Marconi et Siegman, 2017). À ce sujet, les chercheurs Laurence Dierickx et Carl-Gustav Linden, étant d'avis que le terme *intelligence augmentée* est plus approprié, précisent que « la plupart des expériences en matière d'automatisation témoignent non pas d'un remplacement du travail humain, mais bien d'une transformation du travail » (Dierickx, 2021a, p. 7). Nous constatons donc que les outils liés à l'IA s'intègrent en collaboration avec le travail journalistique et dans le but d'augmenter leur pratique, ce que Nicholas Diakopoulos nomme hybridation : où le meilleur de la machine et de l'humain sera mis à contribution.

Par ailleurs, les outils liés à l'IA comptent trois champs d'application lorsqu'il est question de les intégrer dans le domaine journalistique. Tout d'abord, il est possible d'en faire usage dans la collecte et l'analyse d'informations en faisant ressortir des tendances dans les jeux de données permettant de concevoir des postes en journalisme de données ou encore en effectuant des veilles automatiques à partir de signaux faibles sur une exclusivité journalistique (spécialisation), que ce soit le sport, l'économie, la finance, la politique ou autres (Marconi, 2020). Ensuite, certains outils développés s'inscrivent dans la production de nouvelles en facilitant la transcription automatique des entrevues (Journalism AI *et al.*, 2019) ou en automatisant l'écriture de nouvelles routinières dans les sports ou en finance. Finalement, on retrouve des initiatives dans la distribution des nouvelles en permettant d'automatiser les contenus placés en une sur les différentes pages des sites de presse pour personnaliser l'expérience du lecteur (The Globe and Mail Inc., 2020a, 2020b) ou même en mettant de l'avant automatiquement des articles intemporels sur les réseaux sociaux de l'entreprise de presse (Team Twipe, 2021).

2. INTÉGRATION RÉUSSIE DES OUTILS LIÉS À L'IA

Avant d'énumérer certaines intégrations de la technologie sous les trois champs d'application (la collecte et l'analyse d'information, la production et la distribution) dans différentes salles de rédaction aux États-Unis, nous considérons essentiel de présenter deux initiatives. La première est un programme intensif sur l'intégration de l'IA en journalisme du Knight Lab de l'Université Northwestern, un laboratoire interdisciplinaire axé sur le

7. Traduction libre de : « *When journalists use those tools to enhance their reporting, writing and editing, we call it augmented journalism* ».

journalisme et les nouvelles technologies. La deuxième est une recherche de la Knight Foundation, une fondation des frères James L. et John S. Knight finançant des projets en journalisme, concernant 130 initiatives liées à l'IA mises en place dans des salles de rédaction.

2.1 Initiative et constats du Knight Lab et de la Knight Foundation

Tout d'abord, le projet du Knight Lab conçu sous forme de cours intensif de dix semaines illustre deux éléments au cœur de notre recherche, soit le besoin de collaboration et l'implantation d'une équipe multidisciplinaire. Dix équipes regroupant des étudiants en ingénierie, en journalisme et en communication marketing ont développé des prototypes d'outils dans le but d'explorer la façon dont l'informatique, plus précisément l'intelligence artificielle, peut être bénéfique pour le journalisme (Sandalow, 2021). Pour le professeur en journalisme à l'Université Northwestern Jeremy Gilbert, le but de cette initiative est de montrer que «le journalisme est plus fort lorsqu'il est produit en collaboration avec des experts en informatique⁸» (Sandalow, 2021). Pour les projets développés lors de ce programme intensif, une des équipes a conçu un outil qui génère des suggestions dans le but de simplifier la longueur des articles et de les rendre plus facilement accessibles à des personnes possédant un niveau de lecture plus faible. Ce dernier fonctionne en signalant certains mots complexes, en proposant des boîtes d'informations contextuelles, de même que des synonymes plus simples à comprendre. Il n'est toutefois pas indiqué si le produit a été testé par des utilisateurs. Un autre des projets fait usage des techniques d'apprentissage profond – *deep learning* – pour détecter dans le texte du journaliste certains mots à connotation politique qui pourraient nuire à l'impartialité de l'auteur. L'outil propose des solutions de rechange plus neutres. Bien qu'il soit intéressant, ce type d'utilisation comporte aussi certaines limites. Il faut savoir que les programmes d'IA sont formés à partir de données sur un sujet. Dans ce cas-ci, ce sont fort probablement des termes identifiés comme étant partisans ou non. Il y a donc une certaine subjectivité à la base de la conception de l'outil étant donné que ces termes doivent être trouvés et identifiés comme tels. Il faut donc être transparent lors de la mise en place du programme concernant la base de données qui est fournie pour corriger certaines lacunes ou erreurs possibles. Également, plusieurs équipes ont conçu des outils

8. Traduction libre de : « *The real heart of the class is the idea that any journalism is stronger through collaboration with computer science* ».

aux usages variés concentrés principalement sur les médias locaux : l'une d'elles a produit un traqueur visant à déterminer les quartiers sur- ou sous-représentés dans la couverture médiatique. Un autre projet s'est attardé à la façon dont les nouvelles sur un même sujet sont couvertes selon que le média est considéré de gauche ou de droite. Pour le professeur en informatique à l'Université Northwestern Larry Birnbaum, qui a travaillé sur cette initiative, « ce cours n'est que la pointe de l'iceberg quant aux possibilités qu'offre le jumelage du journalisme aux technologies⁹ ». Sommairement, ce cours intensif montre comment la collaboration entre divers groupes d'horizon différents permet la conception d'outils innovants.

Quant à la Knight Foundation, celle-ci a lancé, en mai 2021, une bourse de trois millions de dollars américains pour permettre aux médias locaux de faire usage des outils liés à l'IA. Ce programme a été lancé à la suite d'une recherche analysant 130 projets où des outils liés à l'IA étaient utilisés. Celle-ci constate que les efforts des médias nationaux et internationaux quant à l'intégration des outils liés à l'IA supplantent ceux des médias locaux. Pour les auteurs de l'étude, la raison est simple : « les grosses organisations disposent de plus de ressources humaines, temporelles et financières qu'elles peuvent dédier aux programmes d'innovation et d'expérimentation¹⁰ » (Keefe *et al.*, 2021). Les résultats à notre questionnaire font état d'une tendance similaire dans les usages de la technologie (St-Germain et White, 2021). En effet, nous avons remarqué que les médias utilisant le plus souvent des outils liés à l'IA sont ceux qui disposent de meilleurs moyens financiers. Bien que nous n'ayons pas accès aux états financiers de chacun de nos répondants, nous avons remarqué que seulement le *Globe and Mail* (quotidien national de référence) et La Presse canadienne/The Canadian Press (agence de presse nationale au Canada) en faisaient un usage approfondi. Nous avons aussi remarqué que les plus petits médias expérimentaient ces outils de façon limitée. Le journal *Métro Montréal*, propriété de Métro Média, a précisé par courriel quelques mois après leur réponse à notre questionnaire qu'il faisait usage de la technologie, sans donner plus de détails. Bref, il semble en effet y avoir une inégalité quant à l'accès à ces outils selon la situation financière de l'entreprise, mais également selon l'accès à l'expertise. La Knight Foundation, dans

9. Traduction libre de : « *This class is just the tip of the iceberg at the intersection between media and journalism, and technology* ».

10. Traduction libre de : « *The larger organizations have more resources, in time, people, and money, to devote to innovation and experimentation. They also may have larger upsides for those investments* ».

son communiqué annonçant l'initiative, précise d'ailleurs que les organisations médiatiques locales manquent d'expertise pour mener à terme leur projet utilisant des outils liés à l'IA (Cheung, 2021). De même, la recherche fait remarquer que près de la moitié des outils liés à l'IA conçus pour les médias d'information s'inscrivent dans la catégorie « augmenter les capacités rédactionnelles » (*augmenting reporting capacity*), ce qui s'inscrit dans le champ de la collecte et l'analyse d'information afin d'aider les journalistes à faire leur travail (Keefe *et al.*, 2021). La deuxième catégorie la plus utilisée est ce que les auteurs appellent « réduire les coûts variables » (*reducing variable cost*) comprenant la rédaction automatisée d'articles simples permettant de libérer des journalistes qui effectuaient autrefois la rédaction. Les usages dans la catégorie « optimisation des revenus » (*optimizing revenue*) arrivent en troisième place. Cette section comprend le fait de proposer des articles aux lecteurs selon leurs intérêts ainsi que la gestion des abonnements et du mur payant, comme le fait le *Globe and Mail*.

2.2 Série d'intégration réussie d'outils liés à l'IA aux États-Unis

Comme nous l'avons mentionné précédemment, les outils liés à l'IA en journalisme ont différentes visées selon qu'ils sont utilisés lors de la collecte et l'analyse d'information (1), la production (2) ou la distribution (3). Les prochaines lignes présentent diverses initiatives selon le champ d'application.

Tout d'abord, en ce qui concerne la collecte et l'analyse d'information, on retrouve certains usages pour alerter les journalistes lorsque des situations exceptionnelles se produisent, mais, parfois, la technologie permet de produire de nouveaux types d'articles à partir d'une récolte de données jusque-là difficile à produire. Au cœur de ces outils se trouve la plupart du temps l'apprentissage machine qui a la capacité d'analyser plusieurs milliers de documents et ainsi d'établir des tendances à partir de critères prescrits. *BuzzFeed News* a fait appel à cette technologie pour un article en 2016 pour chercher des avions-espions américains armés d'équipement de surveillance dans le ciel (Aldhous, 2017). Comme base de données principale, le média a utilisé le site de suivi des vols *Flightradar24*. Ensuite, l'équipe de *BuzzFeed News* a analysé plus de quatre mois de données s'échelonnant de la mi-août à la fin de décembre 2015. Par la suite, elle a défini les caractéristiques de vol de plus de 20 000 avions durant cette période : taux de rotation, vitesse, altitude, parcours et durée des vols, en plus d'informations propres à chaque avion : manufacturier, modèle et

numéro à quatre chiffres du transpondeur. L'équipe a ensuite jumelé 500 avions provenant de cette base de données à une autre base comprenant 97 avions identifiés comme appartenant au Federal Bureau of Investigation (FBI) et au Département de la sécurité intérieure des États-Unis et colligée à partir de différentes sources, dont un article de l'Associated Press. Par l'entremise d'un algorithme nommé *random forest*, l'outil a été en mesure d'identifier, selon les paramètres qu'il juge le plus adaptés parmi ceux qui ont été énumérés plus haut, les avions ressemblant aux tendances du FBI et du Département de la sécurité intérieure des États-Unis. Les journalistes admettent cependant que l'algorithme n'était pas infaillible, identifiant par erreur des séances de parachutisme. Ce nouvel algorithme a permis de produire deux articles montrant que des avions-espions américains armés d'équipement de surveillance survolaient certaines villes américaines, en plus d'être présents lors d'une chasse à l'homme au Mexique en lien avec les cartels de drogue du pays.

Un autre exemple fréquemment cité est celui de l'*Atlanta Journal-Constitution* (AJC) ayant révélé, dans une série d'articles, que plusieurs médecins sanctionnés pour abus sexuel exerçaient toujours la profession un peu partout aux États-Unis (Robbins *et al.*, s. d.). Comme l'explique Meredith Broussard dans son livre *Artificial Unintelligence*, un journaliste du AJC avait remarqué que près des deux tiers des docteurs accusés d'abus sexuel envers des patients en Géorgie n'avaient soit jamais perdu leur licence, soit été pardonnés après avoir été sanctionnés (Broussard, 2019, p. 42). Les journalistes du média ont donc voulu pousser la recherche plus loin et établir si la Géorgie faisait figure d'exception. Après plusieurs demandes d'accès à l'information en vain, une cinquantaine de programmes informatiques ont été conçus pour construire une base de données en récoltant les informations à même les sites des régulateurs. Ils ont amassé plus de 100 000 documents. Pour naviguer dans ceux-ci, ils ont utilisé l'apprentissage machine dont le mandat était d'évaluer les probabilités que le dossier soit relié à des abus sexuels. L'outil a permis de diminuer ce nombre à environ 6000 dossiers. Ces derniers ont ensuite été analysés à la main par l'équipe de journalistes.

Les outils liés à l'IA prennent aussi diverses formes lorsqu'ils sont utilisés pour la production de nouvelles. En effet, il est possible d'automatiser l'écriture de courts articles à partir d'un gabarit dont la seule chose qui change, outre quelques variations préprogrammées du vocabulaire, ce sont les données intégrées dans l'article. Un des exemples les plus efficaces pour illustrer cet usage est le Quakebot (L.A. Times, 2019). Travaillant à partir

d'un script informatique récoltant automatiquement les données auprès du U.S. Geological Survey, l'outil répertorie les tremblements de terre enregistrés par l'organisme et produit de courts articles visant à informer les citoyens rapidement après les faits. Chaque article est ensuite révisé. Au besoin, un reporter ajoute des éléments de contexte (L.A. Times, 2019). D'autres outils automatisant la production d'articles existent, comme l'outil Heliograf au *Washington Post* ou Cyborg chez *Bloomberg* (Dans, 2019; Levenson, 2014). Au *Washington Post*, Heliograf a été initialement conçu pour aider les journalistes lors de leur couverture des médailles durant les Jeux olympiques de 2016 à Rio. L'outil a ensuite été réutilisé pour couvrir les élections présidentielles de la même année. En 2020, Heliograf s'est réinventé pour couvrir les résultats électoraux dans un format audio du type podcast (WashPostPR, 2020). Chez *Bloomberg*, Cyborg permet d'extraire les éléments importants des rapports financiers d'entreprise pour générer automatiquement de petits articles (Beckett, 2019, p. 31) et transmettre des alertes aux journalistes. De même, il existe certains usages de la technologie pour automatiser la transcription intégrale d'entrevues. Google Cloud *speech-to-text* (Google, s. d.) est disponible en français. De son côté, l'Agence France-Presse (AFP) compte aussi un outil nommé Transcriber (AFP, s. d.) mais il est utilisé exclusivement à l'interne.

Pour ce qui est de la catégorie distribution de nouvelles, l'usage des données générées par les lecteurs sur les différentes plateformes médiatiques pour créer des profils et proposer du contenu personnalisé a la cote. Outre le *Globe and Mail* qui utilise énormément cette technologie remportant bon nombre de prix à l'international pour leur outil *Sophi* (*sophi.io*/), le *New York Times* et le *Wall Street Journal* ont aussi produit des initiatives similaires (St-Germain et White, 2021; The *Globe and Mail*, 2021). Pour le *New York Times*, l'intégration de ce qu'ils appellent la *curation algorithmique* (agrégation) était nécessaire considérant que le média produit plus de 250 articles quotidiennement (Coenen, 2019). Ce faisant, il devait instaurer une technique permettant d'adapter la distribution des articles aux intérêts des lecteurs. Ils ont donc conçu un algorithme utilisant l'apprentissage machine avec le mandat de jumeler les données de localisation du lecteur à l'article où il a le plus de chance de cliquer. Le média nomme cette stratégie *contextual bandits* : une série d'articles est proposée à différents lecteurs provenant de différents États et l'outil enregistre lorsque l'utilisateur clique ou non sur l'article suggéré. C'est ainsi que la base de données se construit et permet au journal d'optimiser la distribution de son contenu. Cette technique a aussi été intégrée à la

section « Editor's Pick » du site où 30 articles sont choisis par les éditeurs. Bien que ce soit une sélection manuelle, la distribution est gérée par l'outil (Coenen, 2019). Au *Wall Street Journal*, l'outil développé sert à adapter le mur payant du site Internet selon le lecteur et ainsi prédire la probabilité que ce dernier s'abonne au média et consomme le contenu barré (Wang, 2018). Leur système fonctionne avec un score basé sur 60 données différentes, dont le fait que ce soit la première fois ou non que l'utilisateur se présente sur la plateforme, le navigateur et l'appareil utilisé ainsi que les articles sur lesquels il clique. Utilisant aussi l'apprentissage machine, le média est en mesure de coupler les données et de construire des recommandations et des profils de lecteur. Ces profils sont ensuite utilisés pour proposer le contenu et déterminer les accès du lecteur au site. Parmi les usages possibles entre le profil et l'accès au site, le journal est en mesure de proposer des accès temporaires à la plateforme en échange d'une adresse courriel permettant d'augmenter leur base de données et de fidéliser le lecteur en lui proposant des infolettres selon ses intérêts. Par ailleurs, le *Wall Street Journal* certifie que les données récoltées sont toutes anonymisées. Avec ces deux exemples, il est possible de remarquer les possibilités offertes par la collecte de données de première main¹¹ et la distribution des contenus. Il n'en demeure pas moins qu'il y a un risque à personnaliser ainsi le profil des utilisateurs. En effet, et les répondants à notre questionnaire l'ont abordé, cela enferme le lecteur dans une bulle où il n'est plus sujet à des contenus remettant en question ses intérêts, ses opinions et autres. Il est donc nécessaire d'adopter une approche équilibrée pour maximiser la rentabilité des contenus, sans enfermer le lecteur dans une bulle de filtre ou une chambre d'écho.

Ces quelques exemples sont fort intéressants parce qu'ils illustrent bien les capacités des outils liés à l'IA, mais aussi leurs limites. En effet, si ceux-ci sont en mesure d'effectuer à plus grande échelle et plus rapidement une tâche normalement demandée à un journaliste, l'IA doit toutefois être corrigée et supervisée, en plus d'être formée à partir d'une base de données de qualité. Nous y reviendrons en conclusion de ce chapitre. Par ailleurs, il est important de revenir sur certaines craintes formulées par les répondants à notre étude quant à la supervision des outils liés à

11. Traduction libre du mot anglais *first-party data* qui se veut une technique de récolte de données où les informations sont prélevées directement de l'utilisateur par le site Internet visité. Concrètement, cela peut se produire lorsque l'utilisateur s'authentifie sur le site, répond à des questions sur ses préférences ou bien navigue tout simplement sur les pages du site Internet.

l'IA, considérée comme une tâche sans valeur ajoutée, de même que par rapport aux risques quant à l'opacité des algorithmes, ceux d'une possible surabondance de contenus générés par une automatisation des articles, de la possible réduction de la main-d'œuvre ainsi que des erreurs probables et du risque de présenter une information homogénéisée au lecteur en ne lui montrant que les articles qu'il veut voir (St-Germain et White, 2021). Bien que chacune de ces craintes soit légitime, certaines peuvent être réglées facilement. En effet, concernant la supervision comme tâche sans valeur ajoutée, tout dépend de l'usage de l'outil. Par exemple, si ce dernier sert à automatiser un article simple et répétitif dans la salle de nouvelles, comme au *L.A. Times*, cela libère beaucoup de temps au journaliste ou au pupitreur alors que sa tâche n'est maintenant que de relire et mettre l'article en ligne. Il y a donc un avantage à superviser l'outil le temps qu'il se développe et à relire les quelques textes produits. Sur le plan de l'opacité des algorithmes, tout est une question de transparence autant au niveau de la base de données utilisée que de son processus d'analyse et de traitement de ces données. La plupart du temps, on remarque surtout, lors de la collecte et de l'analyse, que les entreprises construisent leur propre base de données et ensuite fournissent des paramètres à l'outil, comme ce fut le cas à l'*Atlanta Journal-Constitution* et à l'International Consortium of Investigative Journalists (un exemple non abordé dans cet article) (Guevara, 2019; Robbins *et al.*, s. d.). Également, c'est à ce moment que les cinq V (volume, *velocity* [fréquence], valeur, variété et véracité; Linden *et al.*, 2019, p. 7), qui servent à analyser les données et qui seront abordés dans la conclusion, prennent tout leur sens. Les craintes quant à la surabondance et à l'homogénéisation des contenus soulèvent de bonnes questions à se poser lors de l'intégration de ces outils. Par exemple, est-ce réellement pertinent de produire plusieurs petits articles répétitifs, n'y a-t-il pas un risque de diminuer la portée potentielle des articles? Comment diversifier le contenu affiché au lecteur tout en personnalisant l'offre? Ce sont des questions légitimes auxquelles les réponses peuvent varier selon le média et la visée souhaitée de l'outil. D'où l'importance de bien connaître le secteur qui sera automatisé, mais aussi la ligne éditoriale du média ainsi que son public cible, pour que l'outil s'inscrive en continuité avec les visées de l'entreprise et non pas en opposition (St-Germain et White, 2021). Il en est de même pour les erreurs potentielles de l'outil. Si le média définit bien sa base de données, les risques sont moindres. Finalement, il est vrai que les répercussions futures des outils liés l'IA sur le domaine journalistique en lien avec une potentielle perte d'emplois

sont difficiles à prévoir. Cependant, les experts semblent plutôt optimistes à cet égard. En effet, le pourcentage de travailleurs à risque de voir leurs postes automatisés varie entre 5,5 % pour les éditeurs et 11 % pour les reporters et les correspondants, selon les estimations de deux chercheurs de l'Université d'Oxford (Frey et Osborne, 2017). De plus, l'objectif est que ce petit pourcentage de journalistes touchés soit déplacé pour accomplir un travail à forte valeur ajoutée, soit en collaborant sur des enquêtes qui autrement n'auraient jamais vu le jour, soit en mettant sur pied des postes de journalistes de solutions et de journalistes de données ou encore des infolettres spécialisées et de longs balados. Le but n'est pas de faire comme MSN Québec et MSN UK, soit de mettre à pied des éditeurs et de les remplacer par un outil automatisé et surtout sans supervision (Proulx, 2020 ; Waterson, 2020a, 2020b). De plus, le domaine est déjà en crise au Canada. Plusieurs pertes d'emplois ont lieu – 3011 de façon permanente ou temporaire en 2020 et 2021 selon les données de l'école de journalisme de l'Université Ryerson (maintenant la Toronto Metropolitan University) et la Canadian Association of Journalists – causées par la perte de revenus publicitaires aux mains des géants du numérique, comme le montre le Centre d'études sur les médias de l'Université Laval, ou bien à cause de la pandémie de COVID-19 (CEM, 2020 ; Wechsler, 2021). Aux États-Unis, 6150 emplois en journalisme ont été perdus depuis mars 2020 selon un décompte de la Columbia Journalism Review (Miller, 2021). L'intégration des outils liés à l'IA a davantage comme objectif de trouver des solutions aux problèmes existants et de venir en aide aux médias en mettant sur pied différents projets utilisant la technologie pour permettre aux journalistes d'effectuer des tâches avec de la valeur ajoutée.

CONCLUSION

Nous avons abordé les nombreuses possibilités qu'offrent les outils liés à l'intelligence artificielle dans le domaine journalistique. Toutefois, les initiatives décrites plus haut ne sont que des exemples. Rien ne garantit que la reproduction de ces dernières soit viable pour toute salle de rédaction. C'est justement l'un des points apportés par le directeur de l'innovation technologique et journalistique de la Knight Foundation, Paul Cheung, qui mentionne que ce n'est pas parce que le *New York Times* utilise des outils liés à l'IA pour filtrer les commentaires que cet usage va fonctionner pour tous les médias (Brennan, 2021). Le conseil est plutôt de connaître ce que son organisation fait de bien et d'adapter l'intégration de ces outils aux visées de l'entreprise.

Finalement, un élément central à la production de ces outils est les données. Ce point est extrêmement important, car, comme le précise Laurence Dierickx, « sans données fiables, exactes et précises, il n'est pas possible d'obtenir une information fiable, exacte et précise » (Dierickx, 2021b). Autrement dit, sans données de qualité, il n'y a pas d'outils de qualité. À ce sujet, l'Association mondiale des journaux regroupe le fait d'analyser son jeu de données sous l'acronyme des cinq V : *volume*, *velocity* (fréquence), *value* (valeur), *variety* (variété), *veracity* (véracité) (Linden *et al.*, 2019, p. 7). Le premier V, soit *volume*, fait référence à la quantité de données alors que *velocity* réfère à la fréquence à laquelle celles-ci sont mises à jour. Généralement, il est préférable d'avoir un grand ensemble de données mises à jour fréquemment. Cependant, dans certains cas, un bas *volume* peut être compensé par une forte fréquence, comme lorsqu'il est question de la température prise à même un thermomètre. Un haut *volume*, mais une basse *fréquence*, peut aussi être acceptable dans le cas des élections par exemple. C'est ici que la *valeur* de la donnée devient pertinente pour juger si le *volume* et la *fréquence* sont suffisants. Autrement dit, il faut qu'au plan journalistique l'usage de ces données soit pertinent à l'atteinte des objectifs de la salle de rédaction. De même, le quatrième V, *variété*, permet de déterminer les formes que peut prendre la donnée ainsi que sa complexité. Autrement dit, où se cache la donnée, comment peut-elle être récoltée et est-ce facile de le faire ? Le cas de base de données comme lors de la pandémie de COVID-19 demeure relativement simple et un bon exemple de *variété*. À l'opposé, mettre en place des outils concevant automatiquement des articles sur les débats politiques est plus complexe, sans toutefois être impossible (Linden *et al.*, 2019, p. 37). Ce point de la *variété* s'inscrit dans la recommandation pour les salles de rédaction désireuses d'intégrer des outils liés à l'IA de ne pas trop ambitionner au départ et de développer progressivement les outils en répondant adéquatement à des besoins plus pressants (Argoub, 2021). Finalement, le dernier V fait référence à la *véracité* des données utilisées. Autrement dit, proviennent-elles de sources fiables et sont-elles véridiques ? Cet élément n'est pas si différent des questions que se posent déjà les salles de rédaction lorsqu'elles décident de couvrir un sujet ou non.

RÉFÉRENCES

- AFP (s. d.), *AFP Transcriber* | *AFP.com*, <<https://www.afp.com/fr/lagence/medialab/afp-transcriber>>.
- Aldhous, Peter, « Here's how *BuzzFeed News* trained a computer to search for hidden spy planes », *BuzzFeed News*, 7 août 2017, <<https://www.buzzfeednews.com/article/peteraldhous/hidden-spy-planes>>.
- Argoub, Sabrina, « How to create an AI strategy for your newsroom », *Polis*, 31 mars 2021, <<https://blogs.lse.ac.uk/polis/2021/03/31/how-to-create-an-ai-strategy-for-your-newsroom/>>.
- Beckett, Charlie, « New powers, new responsibilities, A global survey of journalism and artificial intelligence », *London School of Economics POLIS*, 18 novembre 2019, <<https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/>>.
- Beckett, Charlie, « The future of journalism post-COVID19: Technology, diversity and collaboration », *London School of Economics POLIS*, 12 juin 2020, <<https://blogs.lse.ac.uk/polis/2020/06/12/the-future-of-journalism-post-covid19-technology-diversity-and-collaboration/>>.
- Beckett, Charlie et Mattia Peretti, « 7 things to consider before adopting AI in your news organisation », *#JournalismAI.pdf*, 2019, <<https://drive.google.com/file/d/1LC61h2V1sQLXidtVf2ZVIB61XAgeGC6m/view>>.
- Bizimana, Aimé-Jules et Oumar Kane, « La presse au défi du numérique : une économie politique des médias régionaux au Québec », *Les Cahiers du journalisme – Recherches*, n° 3, 2019.
- Brennan, Christopher, « Can AI help local news figure itself out? », *Medium*, 19 mai 2021, <<https://medium.com/deepnews-ai/can-ai-help-local-news-figure-itself-out-201fef12d9b5>>.
- Broussard, Meredith, *Artificial Unintelligence: How Computers Misunderstand the World*, Cambridge, The MIT Press, 2019.
- Centre d'études sur les médias (CEM), « Publicité », 2020, <<https://www.cem.ulaval.ca/economie/donnees-financieres/publicite/>>.
- Cheung, Paul, « AI for local news: Advancing business sustainability in newsrooms », *Knight Foundation*, 12 mai 2021, <<https://knightfoundation.org/articles/ai-for-local-news-advancing-business-sustainability-in-newsrooms/>>.
- Coenen, Anna, « How The *New York Times* is experimenting with recommendation algorithms », *Medium NYT Open*, 17 octobre 2019, <<https://open.nytimes.com/how-the-new-york-times-is-experimenting-with-recommendation-algorithms-562f78624d26>>.
- Dans, Enrique, Meet Bertie, Heliograf and Cyborg, « The new journalists on the block », *Forbes*, 16 février 2019, <<https://www.forbes.com/sites/enriquedans/2019/02/06/meet-bertie-heliograf-and-cyborg-the-new-journalists-on-the-block/>>.
- Delporte, Christian, « Jules Verne et le journaliste: imaginer l'information du xxe siècle », *Le Temps des médias*, vol. 1, n° 4, 2005, p. 201-213.
- Dierickx, Laurence, « Journalisme algorithmique : un état de l'art de la recherche », *Les carnets du LaPIJ*, n° 2, 2021a, <https://www.academia.edu/46635024/Journalisme_algorithmique_un_%C3%A9tat_de_l_art_de_la_recherche>.
- Dierickx, Laurence, « Intelligence artificielle et journalisme: une course avec les machines », *Equal Times*, 6 avril 2021b, <<https://www.equaltimes.org/intelligence-artificielle-et?lang=fr>>.

- Frey, Carl Benedict et Michael A. Osborne, « The future of employment : How susceptible are jobs to computerisation ? », *Technological Forecasting and Social Change*, vol. 114, 2017, p. 254-280.
- Google, (s. d.), « Speech-to-Text: reconnaissance vocale automatique », *Cloud Speech-to-Text* | *Google Cloud*, <<https://cloud.google.com/speech-to-text>>.
- Guevara, Marina Walker, « We used AI to identify the sex of 340,000 people harmed by medical devices », *International Consortium of Investigative Journalism*, Novembre 2019, <<https://www.icij.org/investigations/implant-files/we-used-ai-to-identify-the-sex-of-340000-people-harmed-by-medical-devices/>>.
- Journalism AI, VRT NWS et The London school of Economics and Political science, « Introduction to machine learning », *Google News Initiative Training Center*, 2019, <<https://newsinitiative.withgoogle.com/resources/trainings/introduction-to-machine-learning/>>.
- Keefe, John, Youyou Zhou et Jeremy B. Merrill, « The present and potential of AI in journalism », *Knight Foundation*, 12 mai 2021, <<https://knightfoundation.org/articles/the-present-and-potential-of-ai-in-journalism/>>.
- Kelly, Tara, « Episode 10 : Conversation with prof Nicholas Diakopoulos (Northwestern University) », *SoundCloud*, juillet 2020, <<https://soundcloud.com/datajournalism/episode-10-conversation-with-prof-nicholas-diakopoulos-northwestern-university>>.
- Knight Center (s. d.), « News algorithms: The impact of automation and AI on journalism – Journalism Courses », *Journalism Courses Knight Center*, <<https://journalismcourses.org/course/news-algorithms-the-impact-of-automation-and-ai-on-journalism/>>.
- Knight Center Courses, « An overview of algorithmic news media », 8 janvier 2019a, <<https://www.youtube.com/watch?v=gtGGh4pRc1Q&feature=youtu.be>>.
- Knight Center Courses, « Computational thinking », 8 janvier 2019b, <<https://www.youtube.com/watch?v=KzL3eRNCPj8&feature=youtu.be>>.
- LA Times, « What is the Quakebot and how does it work ? », *Los Angeles Times*, 24 mai 2019, <<https://www.latimes.com/la-me-quakebot-faq-20190517-story.html>>.
- Levenson, Eric, « L.A Times journalist explains how a bot wrote his earthquake story for him », *The Atlantic*, 17 mars 2014, <<https://www.theatlantic.com/technology/archive/2014/03/earthquake-bot-los-angeles-times/359261/>>.
- Linden, Carl-Gustav, Hanna Tuulonen, Asta Bäck, Nicholas Diakopoulos, Mark Granroth-Wilding, Lauri Haapanen et Hannu Toivonen, « News Automation: The rewards, risks and realities of “machine journalism” », 13 mars 2019, <https://jyx.jyu.fi/bitstream/handle/123456789/67003/2/WAN-IFRA_News_Automation-FINAL.pdf>.
- London School of Economics and Political Science (s. d.), « The JournalismAI Collab », *London School of Economics and Political Science*, <<https://web.archive.org/web/20210204200835/http://www.lse.ac.uk/media-and-communications/polis/JournalismAI/Collab>>.
- Marconi, Francesco, *Newsmakers : Artificial Intelligence and the Future of Journalism*, New York, Columbia University Press, 2020.
- Marconi, Francesco et Alex Siegman, « A day in the life of a journalist in 2027: Reporting meets AI », *Columbia Journalism Review*, 11 avril 2017, <<https://www.cjr.org/innovations/artificial-intelligence-journalism.php>>.
- Miller, Gabby, « More than 6,150 news workers were laid off amidst the COVID-19 pandemic », *Columbia Journalism Review*, 10 décembre 2021, <https://www.cjr.org/tow_center_reports/more-than-6150-news-workers-laid-off.php>.

- Newman, Nic, Richard Fletcher, Anne Schulz, Simge Andı et Rasmus Kleis Nielsen, « Reuters Institute digital news report », *Reuter Institute and University of Oxford*, 2020, <<https://www.digitalnewsreport.org/>>.
- Pavlik, John, « The impact of technology on journalism », *Journalism Studies*, vol. 1, n° 2, 2020, p. 229-237.
- Proulx, Steve, « MSN g  r   par un algorithme : 14 emplois perdus au Qu  bec », *Les   crans*, 9 juillet 2020, <<https://web.archive.org/web/20210519150952/https://lesecrans.ca/au-quebec-aussi-msn-remplace-ses-editeurs-par-un-algorithme/>>.
- Robbins, Danny, Carrie Teegardin, Ariel Hart, Jeff Ernsthausen, Ryon Horne, Richard Watkins et Lois Norder (s. d.), « How the Doctors & Sex Abuse project came about », *The Atlanta Journal-Constitution*, <http://doctors.ajc.com/about_this_investigation/>.
- Sandalow, Brian, « Interdisciplinary teams present new solutions in computational journalism », *Northwestern Engineering*, 3 juin 2021, <<https://www.mccormick.northwestern.edu/news/articles/2021/06/Interdisciplinary-teams-present-new-solutions-in-computational-journalism/index.html>>.
- St-Germain, Nicolas et Patrick White, « Les impacts de l'intelligence artificielle sur les pratiques journalistiques au Canada », *SSRN Electronic Journal*, 2021.
- Team Twipe, « “Today in history” : How AI helps to monetize evergreen archive content at Ouest France », *Twipe*, 8 mars 2021, <<https://www.twipemobile.com/monetising-archive-content-using-ai-at-ouest-france/>>.
- The Globe and Mail, « The *Globe and Mail*'s Sophi.io wins digiday media award », *GlobeNewswire Newsroom*, 28 juin 2021, <<https://www.globenewswire.com/news-release/2021/06/28/2254223/0/en/The-Globe-and-Mail-s-Sophi-io-Wins-Digiday-Media-Award.html>>.
- The Globe and Mail Inc., « The *Globe and Mail*'s Sophi wins best digital news start-up award », *Cision*, 20 mai 2020a, <<https://www.newswire.ca/news-releases/the-globe-and-mail-s-sophi-wins-best-digital-news-start-up-award-849480546.html>>.
- The Globe and Mail Inc., « Sophi.io, du *Globe and Mail*, remporte un prestigieux prix de journalisme en ligne », *Cision*, 20 octobre 2020b, <<https://www.newswire.ca/news-releases/the-globe-and-mail-s-sophi-io-wins-prestigious-online-journalism-award-870779451.html>>.
- The New York Times (s. d.), « The *New York Times* – Research & development », <<https://rd.nytimes.com/>>.
- Turvill, William et Phillip Crawley, « How AI helped *Globe and Mail* reach 170,000 digital subs », *Press Gazette*, section News, 29 avril 2021, <<https://www.pressgazette.co.uk/phillip-crawley-interview-globe-and-mail-canada/>>.
- Universit   d'Helsinki et Reaktor (s. d.), « A free online introduction to artificial intelligence for non-experts », *Elements of AI*, <<https://course.elementsofai.com/>>.
- Wang, Shan, « After years of testing, *The Wall Street Journal* has built a paywall that bends to the individual reader », *Nieman Lab*, 22 f  vrier 2018, <<https://www.niemanlab.org/2018/02/after-years-of-testing-the-wall-street-journal-has-built-a-paywall-that-bends-to-the-individual-reader>>.
- WashPostPR, « The Washington Post to debut AI-powered audio updates for 2020 election results », *Washington Post*, 13 octobre 2020, <<https://www.washingtonpost.com/pr/2020/10/13/washington-post-debut-ai-powered-audio-updates-2020-election-results/>>.

- Waterson, Jim, « Microsoft sacks journalists to replace them with robots », *The Guardian*, 30 mai 2020a, <<http://www.theguardian.com/technology/2020/may/30/microsoft-sacks-journalists-to-replace-them-with-robots>>.
- Waterson, Jim, « Microsoft's robot editor confuses mixed-race Little Mix singers », *The Guardian*, 9 juin 2020b, <<https://www.theguardian.com/technology/2020/jun/09/microsofts-robot-journalist-confused-by-mixed-race-little-mix-singers>>.
- Wechsler, Steph, « A year of mapping media impacts of the pandemic in Canada : COVID-19 Media Impact Map for Canada », *JSource*, 11 mars 2021, <<https://j-source.ca/a-year-of-mapping-media-impacts-of-the-pandemic-in-canada-covid-19-media-impact-map-for-canada-update-march-11-2021/>>.

DES USAGES DES *BIG DATA*, DES ALGORITHMES ET DE L'INTELLIGENCE ARTIFICIELLE PAR LES PLATEFORMES AUDIOVISUELLES

Édouard Germain¹, Justine Dorval² et Éric George³

Depuis le milieu des années 1990, dans un contexte marqué par la numérisation généralisée des processus de création, production, distribution et réception des produits audiovisuels, la libéralisation des lois et des réglementations, la financiarisation et la mondialisation des économies, ainsi que par l'essor de nouvelles pratiques culturelles, les industries canadiennes et québécoises du cinéma et de la télévision sont aux prises avec l'arrivée de nouveaux services d'intermédiation. Souvent mises en place par des industries de la communication (télécommunications, informatique et Internet), ces plateformes numériques proposent des formes inédites d'intermédiation (offre sur le Web, absence de publicité, service garanti, abonnement à faible coût, etc.) qui concurrencent celles des distributeurs « traditionnels », leur but consistant dans la majorité des cas à faciliter ou à renouveler l'accès à des produits préexistants selon des modalités plus ou moins sélectives et prescriptives (Mœglin, 2005). Elles tendent à prendre une place croissante en matière de pratiques culturelles et médiatiques, étant à l'origine de plus en plus de la consommation des œuvres alors que, traditionnellement, ces

-
1. Candidat au doctorat en communication à l'Université du Québec à Montréal (UQAM).
 2. Candidate au doctorat en communication à l'Université du Québec à Montréal (UQAM).
 3. Professeur titulaire à l'École des médias de l'Université du Québec à Montréal (UQAM).

dernières étaient proposées soit dans les salles, du moins dans un premier temps pour les programmes de cinéma, soit sous la forme de grilles de programme sur les chaînes de télévision, que celles-ci soient qualifiées de généralistes ou de spécialisées. Elles contribuent de ce fait à redéfinir, à repenser les processus à la fois de création, de production, de distribution et de consommation dans le secteur des industries culturelles, ce que nous allons aborder dans ce texte à propos du secteur de l'audiovisuel⁴.

Aux fins de cet article, nous avons décidé de nous concentrer, d'une part, sur les pratiques de quatre entreprises ayant développé des activités à une vaste échelle internationale, Netflix, Amazon, Disney et Apple, ainsi que sur celles de trois entreprises de taille plus modeste dont les activités se déploient à l'échelle canadienne, soit Bell (avec son service Crave en anglais et en français), Québecor (avec son service en français Club Illico) et Radio-Canada (avec les services Tou.tv en français et Gem en anglais). Il s'agit dans l'ensemble des cas de nous interroger sur les usages des mégadonnées (*Big Data*), des algorithmes et de l'intelligence artificielle par ces plateformes audiovisuelles. Nous retrouvons ce découpage dans le plan de notre texte. Après quelques éléments d'ordre théorique et méthodologique, nous nous intéressons aux plateformes des quatre entreprises qui déploient leurs activités à une échelle internationale. Nous portons ensuite notre attention aux services des firmes de taille plus modeste offerts à l'échelle canadienne, voire principalement québécoise. Puis, dans un quatrième moment, nous tirons divers enseignements à partir d'une analyse contrastive en mettant en évidence plusieurs éléments qui, à propos de l'IA, renvoient à penser qu'en matière de culture et de médias nous sommes toujours entre continuités et émergences, sans oublier d'éventuelles ruptures.

Précisons immédiatement que l'expression « intelligence artificielle » constitue un syntagme discutable (Julia, 2019) et polysémique. Krivine (2021) propose d'ailleurs de distinguer une IA faible et une IA forte, la première renvoyant à la réalisation d'activités particulières très concrètes qui mobilisent les mégadonnées et les algorithmes, la deuxième suggérant une forme d'émulation du cerveau humain. Or, les pratiques des entreprises qui nous intéressent relèvent plutôt de la première définition, celle de l'IA faible. De plus, si parler d'IA a pour but de qualifier « objectivement » l'état d'avancement de nos sociétés, il s'agit aussi d'un vocable qui a une

4. Dans le cadre d'un projet de recherche financé par le Conseil de recherches en sciences humaines (CRSH).

dimension idéologique forte à l'instar d'autres expressions plus anciennes comme « autoroutes de l'information », « société câblée », « société de l'information » (George, 2007). L'IA renvoie (aussi) à un vaste ensemble de discours qui visent à promouvoir autant les orientations de nombreuses entreprises qui y voient matière à formation de profits élevés, voire considérables dans le cadre des transformations du capitalisme, parfois qualifié de numérique, que les activités de bien d'autres professions, parmi lesquelles celles de l'enseignement et de la recherche (Bourdieu, 1979). Nous allons justement constater ci-dessous dans quelle mesure la notion est mobilisée par les entreprises même dans leurs stratégies de communication.

En termes méthodologiques maintenant, nous avons accordé une place importante à l'étude de la sociohistoire⁵ des entreprises à partir d'une étude documentaire comprenant des sources de diverses origines (des rapports officiels des firmes concernées jusqu'aux travaux de recherche qui leur sont consacrés) en nous intéressant, depuis la création de ces entreprises et de leurs plateformes, aussi bien aux activités effectuées en amont en matière de création et production qu'à celles relevant plutôt de l'aval, concernant la distribution et notamment les rapports avec les consommateurs et consommatrices. Nous avons alors cherché à savoir dans quelle mesure les activités ayant donné lieu à la création de ces entreprises demeuraient dominantes ou non dans les décisions prises plus récemment, notamment en lien avec les enjeux qui nous ont intéressés, à savoir quels rôles les dispositifs socio-techniques et leurs usages relevant de l'IA prennent dans les mutations permanentes de la culture et des médias. Nous avons aussi effectué une analyse des façons dont les productions audiovisuelles sont proposées, notamment à partir des pages d'accueil des plateformes développées par ces entreprises en 2019 et 2020⁶ en nous posant les questions suivantes : quelles sont donc les différentes manières de référencer les contenus proposés ?

-
5. Notre histoire, que celle-ci renvoie à des considérations à long terme ou à une échelle temporelle beaucoup plus courte allant jusqu'à celle de la vie quotidienne, est toujours marquée par les rapports sociaux entretenus entre êtres humains, d'où l'importance de parler de sociohistoire (de sociopolitique, de socioéconomie, etc.).
 6. À noter que, si notre corpus des pages d'accueil renvoie à une temporalité plus courte que celle d'ordre documentaire, il paraît intéressant de préciser que, s'il y a manifestement quelques changements sur ces pages au fil des mois et des années – sans compter le fait que ces pages sont parfois en partie différentes en fonction des usages personnels dus justement à l'utilisation des algorithmes –, il semble aussi y avoir une certaine stabilité. Nous avons déjà fait une étude antérieure de certaines de ces pages d'accueil en 2017 (George et Sénécal, 2018) et nous avons déjà fait des observations proches des plus récentes effectuées en 2019 et 2020. Il serait évidemment intéressant de faire une recherche plus systématique pendant plusieurs années de suite.

Trouvons-nous principalement des situations assez classiques fondées sur des classements traditionnels? Pensons ici à des rubriques faisant référence à des choix reposant soit sur la popularité, soit sur les prix gagnés lors de différents festivals. Ou les propositions effectuées relèvent-elles plutôt de nouvelles façons de faire, comme l'emploi d'algorithmes qui mobilisent les mégadonnées? Il est ici possible de renvoyer à certaines rubriques liées à la formation de profils personnels, les choix proposés reposant sur les goûts de la personne concernée. C'est par exemple sur cette base que Netflix s'est fait connaître à une échelle internationale. Finalement, il s'est agi d'étudier les places prises par les algorithmes et l'«intelligence artificielle» dans les stratégies mises en place par les services étudiés en émettant l'hypothèse que les différences constatées s'expliquent largement en fonction des sociohistoires particulières de ces services et des entreprises qui en sont responsables.

1. LES PLATEFORMES AUDIOVISUELLES INTERNATIONALES ET LEUR APPROPRIATION DES MÉGADONNÉES ET DE L'IA

1.1 Netflix et Amazon entre continuités et émergences

À propos de Netflix, il importe tout d'abord de rappeler que l'introduction des algorithmes parmi les choix à vocation stratégique de l'entreprise n'est pas récente, puisqu'elle remonte au début des années 2000. À l'époque, la compagnie offrait un système de suggestion de contenu basé sur les cotes attribuées par ses abonnés et abonnés en vue de prédire leurs préférences (Mota Drumond, 2016). Le 6 octobre 2006, Netflix a poussé l'expérience plus loin en lançant un concours qui offrait 1 million de dollars américains à la personne ou à l'équipe qui serait en mesure d'établir un système de recommandation capable de prédire le classement des films en étant 10 % plus précis que le système utilisé jusqu'alors par la compagnie, à l'époque, Cinematch (Hallinan et Striphass, 2016). Toutefois, ironie du sort, Netflix n'a finalement jamais utilisé l'algorithme gagnant pour deux raisons. Premièrement, celui-ci fut jugé trop cher à mettre en application (*ibid.*). Deuxièmement, Netflix modifia sa manière de collecter les données avec le développement du visionnement en ligne au détriment de la location de supports matériels (DVD). Cela dit, au-delà des coûts que la firme dut déboursier dans le cadre de la compétition sans en obtenir de retombées directes, celle-ci s'avéra tout de même être une belle occasion de se faire connaître, en plus de se positionner comme une entreprise soutenant «l'innovation». Parler d'algorithme et d'«intelligence artificielle» dans le cadre de la stratégie d'une entreprise, c'est aussi – voire avant tout, dans certains cas – affaire de communication.

Malgré le fait que nous en sachions peu sur les usages des algorithmes de la part de Netflix, l'entreprise tente au moins de les employer de deux manières. Il s'agit premièrement d'aider les abonnées et abonnés à choisir le contenu qu'ils ou elles veulent regarder en leur permettant de faire un « choix éclairé » sans système de recommandation trop complexe (Gomez-Uribe et Hunt, 2015). De façon assez convenue, la mobilisation des mégadonnées et de l'IA renverrait ici à la nécessité de faire des choix parmi une offre pléthorique qui a toujours besoin de se distinguer de celle de ses concurrents. La direction de Netflix a elle-même annoncé que, grâce à la mobilisation des algorithmes et d'autres dispositifs sociotechniques comme l'interface de la page d'accueil, elle contrôlait près de 80 % du contenu consommé sur sa plateforme (Delaporte, 2018). Le seul objectif de l'entreprise consiste ici à maximiser en permanence à la fois l'augmentation du nombre de personnes abonnées et la rétention de celles et ceux qui sont déjà membres du club.

À propos de la page d'accueil, l'analyse permet de conclure que l'offre proposée l'est en fonction de catégorisations souvent traditionnelles, à partir desquelles sont établies des sous-catégories plus pointues. De nombreuses rubriques comme « Séries télé acclamées par la critique », « Top 10 au Canada » (ou « Top 10 en France »), « Films primés », « Séries historiques » et « Documentaires » renvoient à des manières de classer des contenus plutôt classiques. Chloé Delaporte parle ici de « médiation ordinaire » car « la manière dont Netflix propose d'accéder à son catalogue est tout ce qu'il y a de plus classique : elle organise les rayons DVD de la Fnac, ceux de feu les vidéoclubs, la programmation des salles de cinéma, des chaînes de télévision... Cela fait plus d'un siècle que le cinéma existe comme pratique culturelle et cela fait au moins 80 ans que les spectateurs ont le choix entre "drame", "comédie" et "film policier" », ajoute-t-elle (Delaporte, 2018).

Deuxièmement, les algorithmes ne servent pas uniquement à recommander des contenus. Netflix les utilise également pour guider la création de contenus originaux. Un des cas iconiques de cette pratique est celui de la série *House of Cards*, première série « originale » de la compagnie californienne produite en 2013. L'entreprise de Los Gatos affirme avoir imaginé le scénario en analysant les préférences et les usages des 33 millions d'abonnées et abonnés qu'elle avait à l'époque (Bullich, 2018). Une telle étude aurait ainsi permis d'élaborer le scénario, de choisir une distribution, un réalisateur, etc., afin d'obtenir une série à succès (*ibid.*). Il est difficile de contester l'engouement créé par *House of Cards*. Toutefois,

il n'est pas certain que celui-ci soit le résultat des données récoltées, d'autant plus que cette série n'est pas une idée originale, mais une adaptation d'une série britannique à succès des années 1990 (George, 2018). Dès lors, il est possible de se demander dans quelle mesure l'utilisation des algorithmes dans la création de contenu par Netflix n'est pas aussi un argument marketing mobilisé dans les stratégies de communication en plus de constituer une « réelle » innovation (Bullich, 2018). Quoi qu'il en soit, l'utilisation des algorithmes a permis à Netflix de justifier son originalité dans la (co)production de contenus originaux. Et il est possible d'ajouter que l'analyse de l'histoire de l'entreprise rend compte de l'existence d'une certaine « culture numérique » liée à un usage des données déjà ancienne.

Cette « culture numérique » est aussi présente chez son principal concurrent, Amazon. Les algorithmes sont utilisés dans l'ensemble des systèmes de recommandation des produits vendus par le géant du commerce électronique depuis 2002 (Rimaud, 2019). Cette culture est d'ailleurs présente non seulement dans les pratiques, mais au moins autant dans les discours et cela n'est pas surprenant étant donné la prétention du fondateur de l'entreprise, Jeff Bezos, a tourné le dos aux pratiques culturelles traditionnelles. En effet, ce dernier n'a eu de cesse « de répéter qu'il avait peu de respect pour les vieux gardiens des médias, dont le modèle d'affaires remontait à l'âge analogique et dont la fonction était de rédiger des critiques subjectives indiquant au public quoi consommer. Dans le nouvel âge d'Internet, n'importe qui [peut] créer quelque chose, trouver une audience et permettre au marché de déterminer la forme économique adéquate » (*ibid.*). Et les algorithmes ainsi que l'intelligence artificielle seraient les outils au service de ce « nouvel âge ».

De plus, la direction d'Amazon mobilise, ou plutôt mobiliserait, aussi les outils de l'IA susceptibles d'aider la production. Ainsi, après avoir acheté en 2011 la compagnie européenne Lovefilm, elle l'a transformée en « Amazon Studios », une maison de production qui s'appuie sur les données des utilisateurs et utilisatrices pour effectuer des choix⁷. Amazon a alors développé un outil de scénarimage (*storyboarding*) automatisé permettant aux internautes de choisir parmi plusieurs pilotes afin d'orienter la production (*ibid.*). La participation des internautes ne se limite pas à un tel choix puisqu'Amazon procède également à « l'analyse des pauses, du retour en arrière pour revoir

7. Dans un ordre d'idées voisin, Amazon a acheté les studios MGM en 2022 pour 8,5 milliards de dollars US, ce qui lui a permis d'enrichir son catalogue de 4000 films dont ceux de la franchise James Bond.

une scène, le passage rapide sur certaines scènes et l'arrêt complet du visionnage» (George, 2018). La production se fait donc avec pour objectif d'aboutir à la meilleure série possible à l'aide des données de « bêta-testeurs ». C'est d'ailleurs de cette façon qu'Amazon a produit *Alpha House* en 2013, une série qui n'a pourtant pas obtenu le succès escompté (*ibid.*). On retrouve bien ici l'idée selon laquelle le succès d'une production culturelle n'est jamais assuré.

Il importe de bien saisir que, s'il y a des ressemblances notables entre les usages des algorithmes de la part de Netflix et d'Amazon, en revanche, les objectifs diffèrent notablement en fonction des profils des deux entreprises. Alors que, dans le premier cas, les activités audiovisuelles en ligne constituent le cœur des activités de la firme, dans le deuxième cas, celles-ci ne représentent qu'une fraction de l'ensemble des produits proposés. Le service de Prime Video doit être considéré comme une extension du vaste catalogue d'Amazon qui a fait connaître le géant mondial du commerce en ligne. Il est d'ailleurs possible d'effectuer une recherche des vidéos estampillées « Amazon Originals » sur le site d'Amazon, et pas uniquement par la plateforme Prime Video. De plus, l'activité de l'entreprise dans le secteur audiovisuel s'accompagne d'autres développements dans les filières⁸ culturelles de l'édition et de la musique, et bien au-delà étant donné qu'il est aussi possible d'acheter énormément de biens et de services très divers grâce au géant mondial du commerce en ligne. L'étude historique des activités d'Amazon indique d'ailleurs que, si elle a commencé avec un statut de « simple détaillant », elle a en revanche développé très rapidement un vaste ensemble de procédés computationnels chargés de mieux « orienter » les choix de ses toujours plus nombreux consommateurs et consommatrices.

1.2 Disney + et Apple TV + : des utilisations en devenir ?

À la Walt Disney Company, l'usage de sondages et autres méthodes de collecte de données date de bien avant les réseaux socionumériques et autres outils techniques permettant de collecter rapidement de l'information sur son public (O'Flynn, 2019). Toutefois le déploiement de ces

8. La filière peut être définie comme « une suite d'opérations successives articulées autour d'un substrat technique commun ; un certain nombre de secteurs industriels reliés entre eux, en particulier par des échanges de fournisseurs à clients ; l'intensité des relations entre les agents créant des liens d'interdépendance qui sont plus ou moins forts, ce qui se traduit par des intérêts et des contraintes communs » (Ménard, 2014). Elle contribue ainsi à fournir une grille d'analyse des modes de fonctionnements du secteur audiovisuel, composé des filières du cinéma et de la télévision, à la fois en amont (création et production) et en aval (distribution, circulation et consommation).

réseaux à la fin des années 2000 a permis aux entreprises comme Disney d'analyser, de retracer et de cibler leur audience à différents moments de la production et de la distribution. Pour l'entreprise californienne, cette nouvelle possibilité est vue comme une occasion de connaître rapidement les tendances des consommateurs et consommatrices (*ibid.*). Cela permettrait notamment à Disney de prendre moins de risques quand vient le moment de produire du nouveau contenu. Le propos est bien connu.

Néanmoins, en 2016, le président-directeur général de l'entreprise, Bob Iger, a avoué que Disney était en retard dans sa collecte de données concernant la sortie de *Star Wars : l'éveil de la force*, le premier de la nouvelle trilogie (*ibid.*). Ce décalage par rapport aux plateformes comme Netflix s'expliquait notamment par le fait que Disney n'avait pas accès à un support médiatique bien à elle pour collecter puis analyser les données de son audience. L'entreprise avait bien acquis un savoir en matière de goûts de la part de ses consommatrices et consommateurs, mais il s'agissait d'une connaissance avant tout colligée lors des... visites de ses parcs à thèmes ! Un des dispositifs utilisés pour obtenir cette information est en effet le MagicBand, un bracelet permettant de personnaliser l'expérience effectuée dans les parcs que possède la firme. Ce dispositif sociotechnique lui permet de tester de nouvelles idées pour ses films comme *Star Wars* et franchises à l'instar de Marvel avant même d'investir de l'argent dans la production de ceux-ci. La collecte de ces données ne fournit néanmoins qu'un portrait partiel des préférences de son public puisque tous les fans des contenus ne visitent pas systématiquement les « royaumes de Disney ». En conséquence, pour combler cette lacune, la firme investit depuis 2017 dans des logiciels d'apprentissage profond (*deep learning*) qui permettent de retracer les expressions faciales des spectateurs et spectatrices pendant l'écoute d'un film en salle (*ibid.*). À l'aide de l'algorithme « *factorized variational autoencoders* » (FVAE) qui fonctionne par caméras infrarouges, Disney a étudié les réactions faciales d'un public de 150 personnes durant le visionnement de neuf films (Brown, 2017)⁹.

Le lancement de la plateforme Disney+ en novembre 2019 peut être vu comme une manière pour la firme d'avoir enfin accès au même genre de données dont disposent Netflix et Amazon et d'ainsi de guider ses

9. Nos recherches ne permettent pas de conclure que ces innovations ont été mises en pratique à large échelle au moment où nous finissons de rédiger ce texte en 2022. Est-ce lié à la constatation précédente ? Nous n'avons pas non plus connaissance de questionnements d'ordre éthique à ce sujet.

futurs investissements (*ibid.*). Ces nouvelles données couplées à celles qui avaient été amassées dans les parcs à thèmes, les réseaux socionumériques et certaines salles de cinéma permettent dorénavant à Disney d'avoir accès à une quantité d'informations plus importante que jamais sur sa clientèle.

Cela dit, en ce qui concerne sa présentation thématique des contenus à partir de la page d'accueil, une fois de plus, nous pouvons observer des choix assez traditionnels. Dans le cas de la plateforme Disney +, celle-ci se divise entre les sections « accueil », « recherche », « ma liste », « exclusivités », « films » et « séries ». Cette page d'accueil présente un carrousel de dix productions mises de l'avant sur la plateforme de Disney. En dessous, nous retrouvons le logo des cinq grands studios détenus par la Walt Disney Company : Disney, Pixar, Marvel, Star Wars et National Geographic. Cela permet à l'utilisateur et à l'utilisatrice de parcourir le catalogue disponible pour chaque studio. Le reste de la page d'accueil se compose de 21 catégories : « Recommandé pour vous », « Nouveautés », « Continuez de regarder », « Puisque vous avez regardé », « Tendances », « Collections » et autres sous-genres. Ces catégories semblent changer en fonction de la période de l'année. Ainsi, en juillet 2020, Disney mettait de l'avant la catégorie « Célébrons les héros de la communauté noire » probablement en lien avec les revendications du mouvement Black Lives Matters. Mais, encore une fois, nous restons plutôt dans le registre des recommandations effectuées de façon traditionnelle. Seules les catégories « Recommandé pour vous » et « Puisque vous avez regardé » laissent croire que la plateforme offre une forme de contenu personnalisé en fonction de l'utilisation qui en est faite sur la base de procédés qualifiés d'IA.

Venons-en maintenant à Apple. Quelle est donc la politique de l'entreprise en la matière ? Historiquement, celle-ci s'est fait connaître avec l'utilisation d'algorithmes pour suggérer des contenus, que ce soit sur l'App Store ou l'Apple Card. Cette utilisation n'a d'ailleurs pas été sans conséquence pour la firme qui a été accusée de biais sexistes créés par les algorithmes utilisés pour sa carte de crédit (Radio-Canada, 2019). Ceux-ci octroyaient en effet des cotes de crédit inférieures aux femmes par rapport aux hommes qui utilisaient le service de carte d'Apple. Par le passé, elle a également dû modifier le fonctionnement de son algorithme utilisé par son magasin d'application puisqu'il mettait constamment les produits de la « pomme » en valeur. En 2019, Spotify avait d'ailleurs porté plainte auprès de la Commission européenne après que son application fut passée du premier résultat de recherche au huitième en l'espace de quatre ans sur l'App Store (Bazoge, 2019). Cela intervenait au moment où la firme

de Cupertino annonçait la sortie de son service d'écoute de musique en ligne. Ce n'était d'ailleurs pas la première fois que l'App Store faisait l'objet d'une enquête pour des pratiques favorisant ses produits au détriment des autres (*ibid.*).

Afin d'avoir une idée des usages des algorithmes sur Apple TV+, il est sans doute pertinent de porter notre attention à sa plateforme de diffusion en ligne de musique : Apple Music. Si cette dernière utilise les algorithmes afin de suggérer des nouveautés et de bâtir des recommandations basées sur les écoutes antérieures, elle laisse encore une place importante à la recommandation humaine. C'est notamment le cas des listes de lecture créées par les utilisateurs et utilisatrices qu'Apple scrute afin de bâtir ses propres listes, à l'image des pratiques des autres grands services de diffusion numérique de musique, comme Spotify ou Deezer. L'entreprise n'hésite pas non plus à faire appel à des disc-jockeys habitués au monde de la radio pour mettre au point ses listes de lecture. En entrevue avec *The Guardian* en 2015, les dirigeants d'Apple Music décrivaient d'ailleurs l'importance d'avoir un équilibre entre interventions humaines et algorithmes : « *You have to humanise it a bit, because it's a real art to telling you what song comes next. Algorithms can't do it alone. They're very handy, and you can't do something of this scale without 'em, but you need a strong human element* » (Dredge, 2015). Cela dit, il est encore trop tôt pour savoir quel rôle prendront les algorithmes sur Apple TV+. De toute façon, les contenus disponibles sur la plateforme demeurent relativement peu nombreux par rapport à ceux des plateformes abordées précédemment, cette différence s'expliquant par le fait qu'elle accorde la priorité à des œuvres audiovisuelles complètement originales disponibles uniquement sur sa plateforme.

2. DE L'APPROPRIATION DES MÉGADONNÉES ET DE L'IA PAR LES SERVICES LOCAUX

Intéressons-nous maintenant aux services locaux, à savoir Crave, Illico, GEM et Tou.tv. Ainsi que nous allons le constater, ceux-ci ont pour point commun d'accorder la priorité à des modes de référencement et de recommandation des œuvres traditionnelles relevant avant tout de médiations humaines. Ce n'est d'ailleurs pas étonnant à plusieurs égards. D'une part, ces plateformes ont été créées par des entreprises qui sont des acteurs historiques du secteur de l'audiovisuel au Canada et au Québec. D'autre part, elles développent un discours qui met l'accent

sur l'idée selon laquelle leur taille plus petite leur permet justement de mieux saisir les spécificités de leurs publics grâce à une connaissance déjà ancienne des goûts nationaux.

2.1 Crave

Une première analyse permet de constater que la plateforme Crave, développée par l'entreprise numéro 1 au Canada parmi les industries de la communication, Bell, est organisée selon un modèle similaire à ceux de Netflix et d'Amazon. La page d'accueil nous présente un bandeau en carrousel qui met de l'avant des nouveautés (à venir ou récemment ajoutées). À noter que les contenus varient en fonction de la version – francophone ou anglophone – sur laquelle nous nous trouvons. Dessous, la plateforme offre de parcourir le contenu du site Web par entreprise avec laquelle elle est associée – HBO, Showtime et Starz – ou par catégorie – « Kids » et « Collection » en guise d'illustrations. Suit une série de listes classées en fonction des « tendances actuelles », des « coups de cœur » ainsi que de certains genres et thématiques – par exemple « films à succès », « films et séries de chez nous » et « pour les nostalgiques ». Il est également possible de naviguer sur le site Web à l'aide d'une barre au haut de l'écran qui permet d'avoir le choix entre « Séries », « Films », « Mon Crave », « En ondes » et « Crave en cadeau », seul onglet qui ne mène pas à visionner du contenu, mais à offrir une carte cadeau en ligne. Sous les onglets « Séries » et « Films », nous avons une catégorisation classique des contenus en fonction du genre – action, drame, comédies, etc. – de prescriptions plus spécifiques – LGBTQ x Inside Out, 420, etc. – ainsi que de la provenance du contenu – All-Canadian. L'onglet « Mon Crave » donne accès aux options « Reprendre », « Historique d'écoute » et « Ma liste ». Les deux premiers servent à reprendre la lecture de certains titres écoutés récemment ou il y a plusieurs mois et la dernière permet de faire une liste de titres à écouter – option disponible en cliquant sur le bouton « + » situé tout juste à côté du bouton de lecture. Cette section ne fait pas de recommandations particulières en fonction des contenus visionnés ou ajoutés sur « Ma liste ».

Cette analyse conduit à conclure que Crave se range manifestement davantage dans le domaine de la prescription dite « traditionnelle » que dans l'utilisation d'algorithmes pour offrir du contenu personnalisé. À titre d'exemple, la plateforme met davantage de l'avant les catégories comme « *Staff pick* » ou « Aimé par la critique » que les contenus

personnalisés. Elle a d'ailleurs retiré son onglet « *You might like* », le seul susceptible de s'apparenter à du contenu personnalisé.

2.2 Ici Tou.tv. et GEM

La situation est assez proche en ce qui concerne la plateforme Ici Tou.tv. Il importe ici de préciser que celle-ci constitue une émanation de l'organisme de radiotélévision publique, la Société Radio-Canada (SRC). L'offre proposée par Ici Tou.Tv est la même, quelle que soit la personne qui se connecte sur la plateforme, les catégories étant toujours les mêmes. Il demeure l'idée que les œuvres disponibles sur ce service relèvent de la mission de « service public médiatique » de l'organisme (Tremblay, Bizimana et Kane, 2019). Seul élément de personnalisation, avec la création de la partie payante de l'Extra en 2014, a été mis en place *Mon tou.tv*, section du service où il est possible de reprendre un visionnement interrompu, de consulter sa liste de favoris ou bien encore d'avoir accès aux informations liées à son compte. Cette initiative constitue une première tentative de personnalisation du service proposé par la plateforme de diffusion en ligne, mais elle ne mobilise pas d'algorithmes. Nous pouvons tirer les mêmes enseignements de l'étude de la plateforme de la radiotélévision publique en anglais, GEM. Cela dit, s'il n'est jamais fait mention du mot « algorithmes » dans les rapports officiels de la Société d'État, l'intérêt de l'IA est mentionné dans le rapport annuel 2017-2018 à propos de la collecte d'informations sur les habitudes et préférences de visionnement des Canadiens et des Canadiennes sur la plateforme, pour répondre à des « objectifs de découvrabilité, de facilité d'utilisation, et de fidélisation et de mobilisation des auditoires » (Rapport annuel CBC / Radio-Canada, 2018). Par conséquent, les méthodes de proposition de contenu de plus en plus orientées vers les préférences de l'utilisateur et de l'utilisatrice telles qu'elles sont perçues par l'IA pourraient prendre une place croissante par rapport aux modalités traditionnelles de recommandation de l'audiovisuel public.

2.3 Club Illico

La plateforme du groupe Québecor ne semble pas non plus ou fort peu utiliser d'algorithmes pour l'organisation de son contenu et pour cibler les possibles intérêts de ses abonnés et abonnées. Il n'est d'ailleurs jamais mentionné dans les rapports annuels de l'entreprise que sa plateforme a utilisé ou commence à utiliser des outils de mesure sophistiqués

et l'IA afin de cibler les habitudes de consommation ou pour placer son contenu de façon plus stratégique sur sa plateforme¹⁰.

Toutefois, certaines données sont susceptibles d'être utilisées de façon plus modeste, ainsi qu'en témoigne la présentation de contenus listés comme plus populaires dans les catégories défilantes. En date du 15 juillet 2019, la saison 7 de *Journal d'un vampire* (version française de *Vampire Diaries*), série à succès pour jeunes, se retrouvait en début de catégorie pour représenter les sections Horreur, Romance et Suspense. Il existait aussi une section dédiée aux contenus les plus populaires, à l'intérieur de laquelle il était d'ailleurs possible de retrouver la saison 7 de *Journal d'un vampire* à pareille date, ce qui tend à confirmer notre hypothèse.

L'analyse du contenu de la page d'accueil conduit également à considérer que les œuvres mises de l'avant sont avant tout les productions originales de Québecor, contenu destiné en premier lieu au Club Illico ou à une chaîne télévisuelle du réseau TVA, voire aux autres chaînes, spécialisées, du conglomerat québécois. La société mère use depuis de nombreuses années de stratégies de convergence (Crevier et George, 2021) en mettant de l'avant les contenus originaux ou exclusifs à la plateforme, mais aussi toutes les autres productions du groupe Québecor, plutôt que de mettre l'accent uniquement sur les productions qui pourraient être les plus visionnées ou susceptibles de l'être.

3. ÉLÉMENTS DE RÉFLEXION POUR ALLER AU-DELÀ...

Dans cette conclusion, nous allons maintenant nous demander dans quelle mesure, à partir de la prise en compte de nos travaux, il est possible à propos d'IA, de culture et de médias de parler plutôt en matière de continuités, d'émergences, voire éventuellement de ruptures.

Premièrement, il importe de souligner le fait que l'emploi des algorithmes constitue la dernière façon en date de tenter de résoudre un des problèmes traditionnels qu'ont toujours dû affronter les industries de la culture, à savoir qu'il n'est jamais certain qu'une production d'ordre symbolique plaise *a priori* à la personne qui la visionne (Ménard, 2004 ; George, 2014). Chaque œuvre présente certaines particularités en matière de contenu symbolique qui en fait un « prototype » unique. Il est certes

10. L'entreprise québécoise semble avoir recours à ces outils uniquement afin de calculer ses abonnements et comptabiliser ses visionnements.

possible de tenter de profiter du succès d'un film pour essayer de le reproduire en faisant une suite du long métrage, tout comme il est envisageable de proposer un nombre conséquent de saisons dans le cas d'une série « qui marche ». Mais la réussite n'est jamais sûre, que les algorithmes soient mobilisés ou non. Et même si ces derniers s'avéraient assez efficaces pour faciliter le repérage de programmes jugés satisfaisants par les abonnées et abonnés, il serait alors possible de se demander si leur utilisation n'aurait pas tendance à entraîner une diminution du goût pour la nouveauté, la découverte, l'inattendu, les consommateurs et consommatrices s'habituant au cadre rassurant proposé par l'analyse de données de Netflix et autres entreprises (Bullich, 2018). De plus, si changements il y a dans certains traitements des données, en revanche, celles-ci demeurent avant tout mobilisées au stade de la distribution des œuvres. Certes, il y a déjà eu diverses tentatives de mobiliser certaines mégadonnées pour aider à la création de nouveaux contenus, mais ces initiatives n'ont jamais donné lieu à notre connaissance à des résultats spectaculaires. Comme nous l'avons mentionné, les œuvres relevant du domaine symbolique présentent certaines particularités.

Deuxièmement, il importe aussi de souligner le fait que, dans les discours officiels des entreprises, ce ne sont plus des médiations humaines, mais plutôt des médiations techniques, qui sont présentées comme étant à l'origine de la sélection des contenus proposés aux utilisatrices et utilisateurs (Hallinan et Striphas, 2016). Comme nous l'avons rappelé, Jeff Bezos constitue une figure emblématique d'acteur qui développe ce genre de discours orienté vers l'idée de rupture. La réalité s'avère toutefois plus complexe pour deux raisons. D'une part, les médiations techniques dont il est question et qui reposent sur l'utilisation de mégadonnées par des algorithmes sont aussi sociales, car elles passent par la création et la gestion des algorithmes. Jusqu'à preuve du contraire, nous n'en sommes pas à l'ère des machines susceptibles de créer d'autres machines. Mais poser la question du rôle des humains dans la formation des algorithmes et autres processus liés à ladite intelligence artificielle s'avère fort intéressant. D'autre part, les médiations techniques demeurent souvent secondaires par rapport aux médiations humaines en ce qui concerne notamment les systèmes de recommandation ainsi que le montrent nos analyses des pages d'accueil.

Troisièmement, il convient aussi, au sujet de certaines pratiques, de ne pas surestimer systématiquement le rôle des entreprises. L'analyse effectuée par Chloé Delaporte l'a conduite à tempérer le rôle des algorithmes à la suite de leur utilisation par Netflix. Certes, les dispositifs sociotechniques

utilisés pour le calcul des données et le système de recommandation résultant de leur analyse sont des outils particulièrement sophistiqués. Néanmoins, la catégorisation des contenus se fait somme toute de manière plutôt traditionnelle, selon des genres assez génériques (Delaporte, 2018). Nous l'avons aussi constaté. De plus, selon cette autrice, les algorithmes ne conduisent pas systématiquement à conditionner les pratiques de consommation car, sans tomber dans un discours trop enthousiaste quant à l'agentivité des utilisatrices et des utilisateurs, il est possible de les considérer « comme des sujets actifs en capacité de s'approprier les dispositifs de recommandation » notamment en « contournant » ou en « braconnant » les procédés mis en place par Netflix (*ibid.*). Néanmoins, une fois de plus, ce débat n'est pas vraiment nouveau avec, d'un côté, des analyses qui mettent l'accent sur le rôle structurant de l'offre socioéconomique et sociotechnique. Dans le cas présent, il s'agit de mettre de l'avant le dispositif sociotechnique mobilisé sur la plateforme. De l'autre, on trouve d'autres analyses qui, elles, mettent de l'avant les usages des « nouveaux » téléspectateurs et téléspectatrices. Où l'on retrouve ici la pertinence de mobiliser à nouveau Michel de Certeau lorsque ce dernier nous invitait à distinguer production et consommation tout en reconnaissant que cette dernière relevait plus de comportements tacticiens que de comportements de stratégies (Certeau, 1990). Par conséquent, il importe de s'interroger à propos des conclusions de Ted Striplas et Blake Hallinan qui estiment que l'utilisation des algorithmes par Netflix joue sur la nature même des productions culturelles qui sont visionnables et *in fine* ont des conséquences quant à l'ontologie même de la culture (Hallinan et Striplas, 2016). Cette perspective n'est pas à écarter, mais ne s'impose pas d'emblée.

De toute façon, et ce sera le quatrième point de notre partie conclusive, bien que nous sachions quels sont les algorithmes mobilisés par Netflix – principalement six à distinguer les uns des autres – et pour quelles raisons ceux-ci sont utilisés, il est difficile de savoir comment ils fonctionnent. D'une part, parce que les entreprises ne dévoilent à peu près rien de ce qui se passe concrètement entre la collecte des données et la production ou l'acquisition de contenu original. Pour Siobhan O'Flynn, cette absence de transparence fait en sorte que le processus de collecte de données et de recommandation de productions ressemble à celui d'une « boîte noire » (O'Flynn, 2019). D'autre part, parce que, même si nous savons le nombre de catégories qu'elles utilisent – 76 000 dans le cas de Netflix (*ibid.*) – pour classer leurs données en « micro-audiences », il faudrait avoir accès au même genre de technologie pour faire une étude approfondie de

la manière dont ces catégories (ou *tags* en anglais) sont mobilisées pour recommander du contenu.

Or, le fait que les algorithmes et autres composantes de l'IA demeurent aussi obscurs pour la quasi-totalité d'entre nous nous incite à la plus grande prudence avant d'être trop catégorique dans nos réponses. Les recherches doivent se poursuivre... et les compagnies devraient également être ouvertes à l'idée de partager leurs données, ce qui est loin d'être le cas.

RÉFÉRENCES

- Bazoge, Mickaël, « Apple a ajusté l'algorithme de recherche de l'App Store qui favorisait ses propres applications », *iGeneration*, 2019.
- Bourdieu, Pierre, « Les trois états du capital culturel », *Actes de la recherche en sciences sociales*, n° 30, 1979, p. 3-6.
- Brown, Jennings, « Disney Is building facial recognition to figure out when you'll laugh during Toy Story 5 », *Gizmodo*, 2017, <<https://gizmodo.com/disney-is-building-facial-recognition-to-figure-out-when-youll-laugh-during-toy-story-5-1797267294>>.
- Bullich, Vincent, « L'innovation dans les industries culturelles : algorithmes, Big data et prédictibilité », séminaire de doctorat *Médias, nouveaux médias et innovation*, session d'hiver 2018, Montréal, Université du Québec à Montréal (UQAM), 2018.
- CBC / Radio-Canada, *Rapport annuel 2017-2018*, 2018.
- Certeau, Michel de, *L'invention du quotidien. Tome 1 : Arts de faire*, Paris, Folio, coll. Essais, nouvelle édition, 1990.
- Crevier, Julien et Éric George, « Québecor d'hier à demain : d'un conglomérat international à l'ère du papier à une entreprise québécoise à l'ère numérique », Montréal, *Les Cahiers du CRICIS*, 2021.
- Delaporte, Chloé, « Dispositifs innovants, consommation créative ? Netflix ou la recommandation des contenus audiovisuels à l'ère de la prescription algorithmique », Congrès de la SFSIC, *Création, créativité et médiations*, n° 3, 2018, p. 29-38.
- Dredge, Stuart, « Apple Music interview: "Algorithms can't do it alone – you need a human touch" », *The Guardian*, London, 2015.
- George, Éric, « Analyse des discours sur la société de l'information », dans Jean-Michel Ledjou et J. Do Nascimento *et al.* (dir.), *La démocratie à l'épreuve de la société numérique*, Paris, Karthala, 2007, p. 13-32.
- George Éric, « The theory of the cultural industries: A "milieu" for building dynamic knowledge », *Canadian Journal of Communication*, vol. 39, n° 1, 2014, p. 29-54.
- George, Éric, *Note de synthèse*, séminaire de doctorat *Médias, nouveaux médias et innovation*, session d'hiver 2018, Montréal, Université du Québec à Montréal (UQAM), 2018.
- George, Éric et Michel Sénécal, « Les transformations du secteur de l'audiovisuel au Canada et au Québec : place et rôles des plateformes numériques au prisme des pages d'accueil », dans *Création, créativité et médiations*, SFSIC (21^e Congrès), vol. 3: Objets techniques, dispositifs et contenus, 2018, p. 165-176.

- Gomez-Urbe, Carlos A. et Neil Hunt, « The Netflix recommender system : Algorithms, business value, and innovation », *ACM Trans. Manage. Inf. Syst.*, vol. 6, n° 4, Article 13, 2015.
- Hallinan, Blake et Ted Striphas, « Recommended for you : The Netflix Prize and the production of algorithmic culture », *New Media & Society*, vol. 18, n° 1, 2016, p. 117-137.
- Julia, Luc, *L'intelligence artificielle n'existe pas*, Paris, First, 2019.
- Krivine, Hubert, *L'IA peut-elle penser ? Miracle ou mirage*, Bruxelles, De Boeck Supérieur, 2021.
- Ménard, Marc, *Éléments pour une économie des industries culturelles*, Montréal, Société de développement des entreprises culturelles du Québec, 2004.
- Ménard, Marc, « The concept of the industrial channel in the domain of culture, information, and communication : A French speciality? », *Canadian Journal of Communication*, vol. 39, n° 1, 2014, p. 73-88.
- Mœglin, Pierre, *Outils et médias éducatifs : une approche communicationnelle*, Grenoble, Presses universitaires de Grenoble, 2005.
- Mota Drumond, Gabrielle S., *La configuration des usages sur Netflix : le système de recommandation cinematch et la représentation de l'utilisateur* (mémoire de maîtrise), Montréal, UQAM, 2016.
- O'Flynn, Siobhan, « Data science, Disney, and the future of children's entertainment », dans Casie Hermansson et Janet Zepernick (dir.), *The Palgrave Handbook of Children's Film and Television*, Pittsburg, Palgrave Macmillan, 2019.
- Radio-Canada. « L'algorithme de l'Apple Card accusé de sexisme », *Radio-Canada*, Montréal, 2019.
- Rimaud, Mathilde, « Amazon est-il le nouveau géant de la culture et du divertissement? » *Nectart*, vol. 8, n° 1, 2019, p. 67-76.
- Tremblay, Gaëtan, Aimé-Jules Bizimana et Oumar Kane, *Le service public médiatique à l'ère numérique*, Québec, Presses de l'Université du Québec, 2019.

RECOMMANDATION ALGORITHMIQUE ET DILEMMES AGENTIELS

Lily-Cannelle Mathieu¹

Ce chapitre, issu d'une enquête qualitative exploratoire portant sur l'agentivité des personnes usagères de plateformes de diffusion vidéo en continu (*streaming*), traite des dilemmes que peuvent entretenir ces dernières en relation à la recommandation algorithmique de contenus culturels. Plus précisément, le chapitre aborde ces dilemmes agentiels en présentant une classification des facteurs tenus en compte par les personnes usagères alors qu'elles décident, ou non, de poser des actions et inactions agentielles liées au système de recommandation algorithmique.

L'agentivité peut être définie comme la capacité que possèdent les individus d'activer leurs pouvoirs agentiels (Passeron, 2001) sur la base de leurs réflexions (Couldry, 2014), une capacité qui est toutefois enchevêtrée dans une structure sociale ou médiatique contraignant sa forme et ses possibilités (Giddens, 1984 ; Layder, 2006). Le concept des pouvoirs agentiels peut ainsi être opérationnalisé, dans cette étude, comme l'entreprise d'actions (souvent sous la forme de « clics ») générant des données (potentiellement) prises en compte par les algorithmes de recommandation des plateformes fréquentées, ou, alternativement, comme l'inaction volontaire sur une plateforme numérique ayant pour objectif que le système algorithmique n'enregistre pas un « intérêt » ou des données attentionnelles particulières. Dans une telle formule, l'individu disposant d'agentivité est « traité comme

1. Doctorante en anthropologie socioculturelle à l'Université McGill.

régisseur des ressources d'action qui composent sa compétence» (Passeron, 2001, p. 26), comme un agent de l'action.

Il convient également de se pencher sur la définition, ne serait-ce que brièvement, du concept de l'algorithme. En 2014, Gillespie décrit les algorithmes comme étant composés d'une suite d'instructions et de procédures codées qui permettent de transformer, par des calculs mathématiques automatisés, des données entrantes (*input*) en un résultat désiré (*output*) (Gillespie, 2014). Seaver ajoute que les algorithmes ne sont pas des objets stables aux contours bien définis, mais plutôt des « conséquences » d'une variété de pratiques humaines (2017, p. 2), et que les algorithmes n'existent plus au singulier, mais qu'ils sont désormais constitués de multitudes de « sous-algorithmes » analysant chacun différents signaux (Seaver, 2019). Les algorithmes sont continuellement raffinés – par des ingénieurs ou ingénieures ou, plus souvent, par les algorithmes eux-mêmes, lorsqu'il s'agit d'algorithmes récurrents fonctionnant par l'apprentissage-machine (Beer, 2013) – afin que leur performance, qui est mesurée selon des critères établis par des humains, soit maximisée (Beer, 2013 ; Lanier, 2018 ; Seaver, 2019). Il est aussi intéressant de noter que, bien que les algorithmes s'immiscent depuis des années déjà dans le quotidien de plusieurs par le biais des technologies numériques, il semble qu'ils n'aient investi le discours public que récemment, de manière sans cesse plus marquée depuis l'affaire *Cambridge Analytica* des années 2014-2018 (Ovadya, 2018²), et que le concept semble aujourd'hui bien établi dans l'imaginaire social, même s'il n'y est pas bien défini.

Ce chapitre s'intéresse à un système algorithmique particulier, soit la recommandation algorithmique, qui est définie par Ménard (2014) comme un système de filtrage d'informations ayant la fonction de proposer des contenus et des éléments d'informations susceptibles de susciter l'intérêt des usagers et usagères. Comme les autres systèmes algorithmiques, celui-ci se nourrit de grandes quantités de données, qu'il met en relation mathématique (par des statistiques, des comparaisons et des corrélations, par exemple) afin de produire les résultats désirés par ses concepteurs et

2. L'affaire *Cambridge Analytica* a trait à l'exploitation des données personnelles d'utilisateurs et utilisatrices du média socionumérique Facebook par la société Cambridge Analytica, dans l'objectif notamment d'influencer les choix électoraux de ces utilisateurs et utilisatrices en leur présentant des contenus particuliers grâce à un système algorithmique de classement de contenus. Révélée dans les journaux, cette exploitation algorithmique des données personnelles a fait scandale et a contribué à une plus grande présence du concept de l'algorithme dans le discours public (Ovadya, 2018).

conceptrices – dans ce cas, capter l’attention des personnes usagères (Beer, 2013 ; Ménard, 2014) et influencer leur comportement (Lawrence, 2015 ; Kalpokas, 2019 ; Seaver, 2019).

Louise Amooore (citée dans Beer, 2013) soutient que la sélectivité et la direction sont le domaine des algorithmes de recommandation, qui masquent certaines informations ou certains contenus pour en mettre d’autres en évidence sur la base de prédictions – des prédictions qui ne seraient, selon l’informaticien de la Silicon Valley Jaron Lanier, qu’« à peine fiables », mais qui travaillent sans relâche, par l’exploration et l’apprentissage-machine, afin de maximiser l’usage et d’ainsi obtenir toujours plus de données, dont l’accumulation permet aux algorithmes de se raffiner et de produire des résultats sans cesse plus concluants, c’est-à-dire un engagement ou une dépendance toujours plus exacerbée chez les usagers et usagères, dont l’attention constitue le capital principal des plateformes (Pedrycz *et al.*, 2013 ; Lanier, 2018). Ainsi, les algorithmes ne détermineraient pas seulement les sortes d’informations auxquelles les personnes usagères des plateformes numériques sont exposées (Hutchison, 2017), mais, en établissant des régimes informationnels supposant que leur subjectivité est calculable, rendraient ces individus influençables en chorégraphiant de manière pointilleuse leurs expériences de navigation (Introna, 2016). Certains décrivent alors les algorithmes comme des outils de modification de comportement (Lanier, 2018 ; Kalpokas, 2019) conçus dans l’objectif de saisir, d’analyser et de *réajuster* des comportements humains de manière à servir des buts commerciaux bien précis (Lawrence, 2015), comme des entités excédant leur statut d’outil pour tendre vers la manipulation et la semblance d’une prédétermination des choix, expériences et préférences relevant du possible (Arielli, 2018).

Une littérature propre aux algorithmes de recommandation culturelle prend les mêmes refrains plutôt déshumanisants : les algorithmes de recommandation culturelle y sont décrits comme des mécanismes qui *valorisent* et *évaluent* les contenus distribués par les plateformes (Gillespie, 2016, p. 53), se conférant ainsi la tâche d’attribuer de la valeur culturelle à certains contenus plutôt qu’à d’autres, prétendant ainsi que des attributs à caractère culturel peuvent être *quantifiés* (Finn, 2017 ; Beer, 2013). Une telle posture quantitativiste remet selon Hallinan et Striphos (2016, p. 122) l’autorité de la critique culturelle en question et réduit par le fait même la valeur du produit culturel, que les algorithmes identifient comme un agrégat de données plutôt que comme une influence potentielle dans la trajectoire du développement humain. Wright (2015, p. 154) écrit d’ailleurs

qu'en contexte d'abondance culturelle, les systèmes de recommandation algorithmique font désormais partie de l'architecture infrastructurelle et institutionnelle du goût contemporain ; qu'elles influencent les expériences quotidiennes de formation du goût. Beer (2013) explique additionnellement que le processus de recommandation algorithmique établit les conditions de possibilité de nos rencontres avec des contenus culturels, et qu'il excède ainsi l'objectif de prédiction de goût pour contribuer à *former* ou même *prescrire* le goût en bricolant ses horizons. Les recommandations algorithmiques de contenus culturels, qui sont qualifiées par Arielli (2018) d'outils de manipulation culturelle, constituent ainsi les contours de l'expérience culturelle (Beer, 2013) et réduisent la possibilité que les personnes usagères soient exposées à des contenus culturels radicalement différents de ceux qui sont enregistrés dans leurs historiques d'usage (Wright, 2015). Lawrence (2015) soulève par ailleurs la possibilité que la recommandation algorithmique contribue à la dégradation de l'autonomie esthétique des personnes usagères de plateformes numériques étant donné qu'elle cultive chez elles des habitudes de consommation « passive » permettant aux recommandations de s'auto-actualiser, c'est-à-dire de former chez ces personnes usagères des préférences qu'elles auraient elles-mêmes instillées plutôt que de laisser ces individus développer leurs propres goûts et appréciations esthétiques des œuvres.

L'enquête dont il est question dans ce chapitre avait pour objectif de remettre en question ce contrôle supposément aliénant qu'auraient les algorithmes de recommandation sur la consommation culturelle des usagers et usagères de plateformes numériques. En se concentrant sur les personnes usagères des plateformes numériques et sur leur expérience face aux recommandations culturelles algorithmiques plutôt qu'en étudiant les plateformes, leurs configurations et leurs « stratégies techno-commerciales oppressives » – qui obnubilent la plupart des auteurs et auteures d'études sur les données sensibles (*critical data studies*) (Kennedy *et al.*, 2015) –, cette enquête visait à explorer, et possiblement rétablir, la possibilité qu'ont les personnes usagères d'être les agentes de leurs expériences culturelles numériques.

1. DÉMARCHE MÉTHODOLOGIQUE

Cette étude exploratoire analyse le problème de la recommandation culturelle algorithmique au moyen d'une enquête portant sur la relation que de jeunes adultes québécois et québécoises entretiennent avec les

algorithmes de recommandation tels qu'ils se matérialisent sur les plateformes numériques proposant des œuvres audiovisuelles à leurs publics. L'enquête, qualitative, visait à coconstruire des données substantielles (des *thick data* – Latzko-Toth *et al.*, 2017) avec des sujets plutôt que de représenter ces derniers par des jeux de données statistiques – rendant ainsi justice aux perceptions, réflexions et motivations complexes cachées derrière les clics qu'effectuent, à travers des écrans, ces individus dotés de pouvoir agentiel que sont les usagers et usagères (Coûtant et Domenget, 2020).

Dans l'intérêt de garder la dimension vernaculaire du discours que tiennent les personnes usagères à propos de leur relation aux algorithmes de recommandation des plateformes de diffusion en continu, de représenter, tels qu'elles les expriment, les objectifs et motivations alimentant leurs actions (Rollet *et al.*, 2017), et en connaissance du fait que l'expérience de la personne usagère « dépasse largement son discours sur celle-ci » (Savoie-Zajc, 2016, p. 358), un certain « bricolage méthodologique » (Jouët et Le Caroff, 2016, p. 156) visant à enrichir la méthode de l'entretien sociologique était de mise. L'enquête a ainsi été construite en deux étapes distinctes, toutes deux réalisées sur la plateforme de visioconférence Zoom en raison des contraintes sanitaires qui étaient alors en vigueur.

La première étape était une « visite commentée » individuelle (Gallant *et al.*, 2020) lors de laquelle la personne participante partageait son écran d'ordinateur avec la chercheuse afin que toutes deux voient l'interface des trois ou quatre plateformes de diffusion vidéo en continu que la personne participante fréquente le plus souvent. Les plateformes en question se sont révélées être les suivantes : YouTube, Netflix, Prime Video, Facebook Watch, Tou.tv et Noovo. En « visitant » trois ou quatre de ces plateformes, l'une à la suite de l'autre, la personne participante était invitée à commenter l'interface, à naviguer sur la plateforme et à formuler des réflexions ayant trait aux recommandations algorithmiques apparaissant à l'écran, puis à répondre à des questions formulées par la chercheuse. La deuxième étape de l'enquête était un entretien semi-dirigé individuel de format plus traditionnel réalisé une semaine après la tenue de la visite commentée.

La visite commentée et l'entretien, qui ont eu lieu dans une dynamique de coconstruction de sens et d'émancipation du pouvoir réflexif des personnes participantes au cours de l'enquête (Savoie-Zajc, 2016), se sont révélés être d'une grande richesse, mais ont aussi comporté certaines limites, notamment celle que les résultats de l'enquête sont « socialement construits » (Savoie-Zajc, 2016, p. 339) – qu'ils ont été construits et

constitués lors de l'enquête plutôt que de représenter ce qui y est « réellement » extérieur, ou ce qui lui était antérieur. Effectivement, il importe de remarquer que les thèmes abordés lors de l'enquête ne sont pas nécessairement des thèmes auxquels les personnes participantes avaient déjà réfléchi et avec lesquels elles entretenaient précédemment une relation réflexive explicite. Comme l'explique Bellavance (2008), la situation d'entretien, en « forçant » les personnes participantes à s'expliquer, illustre des possibilités qui sont « entièrement réelles », mais qui n'en demeurent pas moins l'expression de moments transitoires résultant d'un jeu – ou d'un combat, peut-être – entre les limites symboliques établies autant par la chercheuse que par les personnes participantes. Gubrium et Holstein (2012, p. 32) notent par ailleurs que l'entretien est « *an occasion for constructing accounts* », une situation qui « crée » des discours qui sont inévitablement liés à leur contexte de création. Un participant a d'ailleurs explicité cette réalité en évoquant à la fin de son deuxième entretien :

Peut-être que ma pensée [n']était pas si évoluée que ça au moment que tu [la chercheuse] posais les questions... T'sais, tu m'as enclenché des réflexions... les réflexions [qu'il avait pu avoir avant de participer à cette étude] [n']étaient pas abouties.

Il convient donc d'appréhender les résultats de cette étude et le discours des personnes participantes comme situés dans un parcours réflexif évolutif chez ces dernières ; situés dans leur processus d'émancipation au sujet de la recommandation algorithmique, ce dernier ayant été provoqué par leur participation à cette étude.

L'échantillon de personnes participantes, qui a suivi une logique « d'étude de cas » plutôt que de « représentativité sociale » (Small, 2009), était intentionnel plutôt que probabiliste (Savoie-Zajc, 2016) et est composé de quatre jeunes personnes adultes (25-35 ans) québécoises (trois jeunes femmes et un jeune homme) recrutées dans le réseau de connaissances de la chercheuse. Au sein de ce petit groupe, on retrouve une certaine diversité d'orientations sexuelles, de bagages ethnoculturels, de statuts conjugaux, de statuts occupationnels, de domaines professionnels ou d'études et de volume de consommation hebdomadaire de contenus audiovisuels.

2. CLASSIFICATION PROPOSÉE : LES COMPOSANTES DES DILEMMES AGENTIELS FACE À UN SYSTÈME DE RECOMMANDATION CULTURELLE ALGORITHMIQUE

Les résultats d'enquête illustrent des propositions issues de travaux antérieurs voulant que la recommandation algorithmique de contenus culturels crée un espace de tension (Karakayali *et al.*, 2018), au sein duquel pratiques, agentivités et algorithmes s'entremêlent (Gillespie, 2014), ainsi qu'un espace structurant, constituant des possibilités tout en les contraignant (Seaver, 2019 ; Siles *et al.*, 2019). Plus que d'y trouver des publics algorithmisés, déléguant tous leurs choix au système algorithmique (Karpik, 2007) ou, au contraire, autonomes, en plein contrôle d'une expérience culturelle numérique modulée selon le contexte attentionnel désiré (Beuscart *et al.*, 2019), on peut y trouver des publics aux prises avec des dilemmes agentiels. Les pratiques culturelles numériques de ces publics sont issues de constantes négociations entre ce qui leur est proposé par les plateformes et leurs algorithmes et ce qu'ils désirent en faire (Mota Drumond, 2020), en tant qu'agents et agentes disposant de certaines ressources d'action.

Les résultats de l'enquête sont proposés sous la forme d'une classification des composantes de ces dilemmes agentiels qu'entretiennent des usagers et usagères, soit (1) des facteurs favorisant leur action ou inaction agentielle liée à la recommandation algorithmique, et (2) des facteurs freinant leur action ou inaction agentielle en lien avec la recommandation algorithmique. À la lumière des résultats, il est proposé que, pour les personnes usagères de plateformes numériques, la décision de poser une (in)action agentielle face aux algorithmes de recommandation culturelle est influencée par différents facteurs, dont la rencontre crée des dilemmes agentiels internes chez ces individus.

De manière générale, sur les plateformes de diffusion vidéo en continu, les pouvoirs agentiels consistent en la gamme des formes de contributions, préformatées par la technique des plateformes, qui font ou peuvent faire l'objet « d'une appropriation individualisée qui témoigne de la subjectivité de chaque participant » (Jouët et Le Caroff, 2016, p. 162). Kitchin et Dodge (2011) réfèrent à ces ensembles de contributions rendues possibles par la structure technologique des plateformes comme des « grammaires d'action codées³ ». Sur les plateformes à l'étude, les pouvoirs agentiels conférés aux

3. Traduction libre par l'auteur.

usagères et usagers par les plateformes varient : il peut par exemple s'agir de cliquer sur un pouce levé ou baissé ou sur un autre symbole tel qu'un cœur, d'écrire un commentaire, d'ajouter un contenu à une liste de lecture, de s'abonner à une chaîne, de cliquer sur « pas pour moi », « masquer ce contenu » ou « ne pas recommander ce type de contenu », de signaler des contenus en cliquant sur un bouton prévu à cet effet, de faire une saisie de recherche libre plutôt que de cliquer sur des contenus recommandés, de cliquer ou non sur des contenus, de passer ou non du temps devant des vidéos et de dérouler ou non l'écran rapidement, de modifier les réglages ou les paramètres de son compte, de quitter la plateforme pour aller en consulter une autre ou s'adonner à une autre activité, ou même de se désabonner à une plateforme. Ces (in)actions, dont la liste proposée n'est pas exhaustive et dont les possibilités varient d'une plateforme à l'autre et au fil du temps, peuvent, selon la complexité des systèmes algorithmiques des différentes plateformes, toutes être prises en compte dans les calculs de recommandation algorithmique, et c'est la décision de les poser ou non afin d'affirmer une agentivité envers le système algorithmique qui constitue le dilemme agentiel dont il est ici question.

3. FACTEURS FAVORISANT L'(IN)ACTION AGENTIELLE LIÉE À LA RECOMMANDATION CULTURELLE ALGORITHMIQUE

3.1 La littératie numérique

L'enquête illustre que la littératie numérique, qui peut être définie comme l'aptitude à *utiliser* et à *comprendre* les technologies numériques (OCDE, 2000), est un des facteurs favorisant l'(in)action agentielle en rapport avec la recommandation algorithmique. Alors que le mot *utiliser* fait référence à des compétences techniques – à une capacité à utiliser les différentes fonctions des plateformes, dans le cas à l'étude –, le mot *comprendre* fait pour sa part référence à une compréhension du fonctionnement d'une technologie – à un certain niveau de compréhension de ce qu'est et comment fonctionne la recommandation algorithmique, dans ce cas.

Tout d'abord, pour pouvoir activer un pouvoir agentiel, il faut savoir qu'on dispose de ce pouvoir ; dans le cas à l'étude, il s'agit par exemple de savoir qu'on peut créer une liste d'écoute, ou, qu'en glissant le curseur sur un certain symbole, on peut avoir accès à des options telles que « masquer » ou « ne pas recommander » certains contenus. À titre d'exemple, bien que les personnes ayant participé à l'enquête disposent toutes d'assez bonnes

compétences d'utilisation des plateformes, elles ne connaissaient pas tous les répertoires d'actions que chacune des plateformes qu'elles fréquentent régulièrement leur offre. Une participante, qui ne savait pas que YouTube lui offrait la possibilité de masquer des contenus, a affirmé, durant l'enquête, que maintenant qu'elle était consciente de cette possibilité d'action, c'est quelque chose qu'elle allait « certainement » prendre en compte les prochaines fois qu'elle naviguerait sur la plateforme. Une autre encore soulignait, à la fin de son deuxième entretien, que le fait d'être consciente qu'en tant qu'usagère, elle dispose de pouvoirs d'action, lui permettra « de les utiliser ». Ainsi, sans trop de surprise, on comprend que ce type de littératie numérique – la capacité d'utilisation – est une composante constitutive du pouvoir agentiel sur les plateformes numériques (Kennedy *et al.*, 2015).

Par la suite, le fait de *comprendre* les principes de la recommandation algorithmique des plateformes culturelles numériques, ne serait-ce que minimalement, favorise l'activation des pouvoirs d'action qui y sont reliés (Claes *et al.*, 2021). De savoir que les plateformes recueillent les données d'usage (ces données issues des comportements numériques) et que ces données sont utilisées dans des calculs algorithmiques sur lesquels sont basées les recommandations permet aux personnes usagères de « produire » ces données d'usage de manière informée. Cela leur permet aussi d'être des génératrices de données actives plutôt que passives, qui ne posent pas que des actions « brutes » (comme simplement peser sur un contenu recommandé ou mettre une mention « J'aime »), mais des actions basées sur la réflexion et une certaine conscience des conséquences de ces actions (Kennedy *et al.*, 2015).

Ainsi, lorsque les personnes participant à l'enquête disent « *je sais que... un geste que je vais poser, ça va l'enregistrer comme de l'information pour, comme, apprendre à mieux me connaître* », « *mes clics sont comptés* », « *les algorithmes de YouTube, je sais comment les activer... dans un sens qui va me convenir, mettons* », « *j'essaie de manipuler l'algorithme de Netflix à mon avantage* », ou même « *tu vois, j'ai bien travaillé mes algorithmes* », elles illustrent qu'une littératie numérique impliquant une compréhension de base du système de recommandation algorithmique est un facteur favorisant l'activation de leurs pouvoirs d'action visant une interaction agentielle avec la recommandation algorithmique.

3.2 L'activation du pouvoir réflexif et de la pensée critique

En plus d'avoir une compréhension minimale du système de recommandation algorithmique, certaines personnes participantes font preuve de pensée critique et réfléchissent activement aux implications et aux conséquences que la recommandation algorithmique peut avoir sur leur développement personnel, sur les producteurs et productrices de biens et contenus culturels, sur des personnes usagères particulièrement influençables, ou même sur la société, de manière générale.

L'activation de ce pouvoir réflexif, à laquelle Kennedy et ses collaborateurs (2015) réfèrent comme une agentivité citoyenne reliée aux structures de donnéification (*datafication*), est une étape de conscientisation qui se construit sur les bases d'une littératie numérique de *compréhension* de ce qu'est la recommandation algorithmique. Il est ici question d'une disposition à la réflexion ou à l'intellectualisation, qui, quand elle est actualisée, nourrit un désir citoyen d'activer les pouvoirs d'(in)action agentielle offerts par les plateformes (Greffet et Wojcik, 2014).

Lorsque les individus activent leur pouvoir réflexif et développent une pensée critique à propos d'un objet tel que la recommandation culturelle algorithmique, leur désir d'activer leurs pouvoirs d'(in)action agentielle avec discernement, de manière intentionnelle ou stratégique – ou même politique, dans certains cas – semble être particulièrement habilité. L'activation d'une pensée critique à propos des algorithmes de recommandation culturelle devient ainsi la pierre d'assise d'un engagement social et citoyen, et devient une habileté civique qui, selon certains (Ylipulli et Luusua, 2019; Ovadya, 2018), pourrait bien être essentielle au bien-être de la société du futur.

3.3 La cybercitoyenneté

En 2014, Greffet et Wojcik remarquaient que « [l]e déploiement des technologies numériques s'accompagne de pratiques sociales et politiques sinon inédites, du moins renouvelées » (p. 127) et qu'il a donné naissance à ce à quoi la littérature scientifique réfère parfois, depuis le début des années 2000, comme la « cyberdémocratie » et la « cybercitoyenneté ». Les définitions d'une telle modalité démocratique sont multiples (Greffet et Wojcik, 2014), mais s'apparentent toutes, d'une certaine manière, à la définition que Poirier (2017, p. 158) relève de la « citoyenneté », c'est-à-dire une « appartenance publique à une communauté politique donnée » à laquelle sont rattachés des droits, des responsabilités et la reconnaissance

d'une « égalité formelle » entre les personnes citoyennes. Cette appartenance publique à une communauté, ces droits et responsabilités, et cette égalité entre les individus, on peut tous les retrouver, sous une forme ou une autre, sur le Web et, plus précisément encore, quoique de manière peut-être plus discrète, sur les plateformes de diffusion vidéo en continu, au sein desquelles les personnes usagères peuvent poser des actions politiques non conventionnelles, relevant de répertoires, vecteurs et cibles nouvelles et diversifiées (Rosanvallon, 2006).

Les résultats de l'enquête illustrent que plusieurs des personnes participantes tendent en effet à percevoir le Web comme un espace au sein duquel elles peuvent s'associer à d'autres individus et appartenir à une communauté. C'est d'ailleurs en raison de ce sentiment d'appartenance à une communauté que toutes les personnes ayant participé à cette enquête perçoivent qu'elles ont certaines responsabilités lorsqu'elles naviguent sur différentes plateformes numériques. Elles activent à cet effet leur pouvoir de « signaler » des contenus ou de les « critiquer » avec un pouce pointant vers le bas lorsqu'elles jugent que ceux-ci sont « inacceptables », que ce soit parce qu'ils contiennent du contenu haineux, sexiste, violent, ou promouvant des idéologies auxquelles elles n'adhèrent pas (tel que le culte extrême de la minceur, par exemple). Ainsi, les personnes participantes ont expliqué n'activer leurs « pouvoirs » à connotation plus négative (le pouce pointant vers le bas ou le signalement) qu'en relation avec des contenus « *qui [les] dérangent vraiment* » ; « *c'est comme un statement politique* » pour exprimer une opinion et influencer le système algorithmique, explique l'une des personnes participantes. On retrouve en elles le ou la *monitorial citizen* de Schudson (1998), qui n'est pas nécessairement actif politiquement, mais qui, dans un mode « davantage défensif que pro-actif, [...] se tient prêt à agir s'il [ou elle] l'estime nécessaire » (Greffet et Wojcik, 2014, p. 132).

Les résultats de l'enquête font aussi ressortir l'observation que les personnes participantes n'utilisent pas leurs pouvoirs d'(in)action agentielle à connotation négative que pour dénoncer des contenus qu'elles jugent socialement inacceptables, et que certaines d'entre elles activent aussi leurs pouvoirs d'(in)action agentielle pour critiquer ou, en quelque sorte, émettre une certaine opinion sur la *modalité de recommandation algorithmique* elle-même. Certaines personnes participantes soulignent par exemple qu'en activant leur pouvoir de *ne pas cliquer* sur des contenus recommandés et d'effectuer des saisies de recherche au lieu de se fier à la recommandation algorithmique, elles peuvent signaler une certaine opinion politique. Elles adoptent ainsi la posture du ou de la *consumer netizen* de Baringhorst et

Yang (2013), pour qui les clics et le temps passé sur une plateforme ou devant des contenus – temps et clics qui constituent le capital principal des plateformes aux vues de leur modèle d'affaires – représentent des votes (Frischmann, 2018) ou, du moins, des modalités discursives d'expression du mécontentement envers un système techno-social jugé oppressif (Greffet et Wojcik, 2014).

À coup de clics, de signalements, de saisies de recherche, de pouces pointant vers le bas ou de déroulements d'écran rapide (signalant volontairement un « désintérêt ») – d'(in)actions agentiellles, en bref –, les personnes usagères ont la capacité d'emprunter un rôle actif qui leur permet de répondre (*talk back*) (Roberge, 2011) au système algorithmique, et ainsi d'établir entre elles et le système de recommandation algorithmique cette réciprocité, ancrée dans la possibilité de formuler des jugements, que concéderait le statut de citoyen ou citoyenne (Lamoureux, 1994 ; Poirier, 2017).

Une participante remarque toutefois que, de manière générale, même si elle adopte une position critique à propos de la recommandation algorithmique, elle a tout de même le sentiment d'en tirer des bénéfices individuels, ce qui fait qu'elle n'active pas nécessairement, ou peu souvent, ses pouvoirs d'(in)action agentielle pour « critiquer » le système de recommandation algorithmique. Elle explique d'ailleurs qu'elle est aux prises avec un certain conflit, qu'elle dit être de nature politique :

Au fil de l'entrevue, en parlant et en élaborant mes pensées, je me rends compte de l'aspect politique des plateformes... Dans le sens où l'espèce de confrontation entre moi, en tant qu'individu, j'ai envie qu'on me propose du contenu pour moi... T'sais, c'est une posture très individualiste et de droite, quasiment, pis après la posture qui est plus dans la collectivité, qui est plus démocratique, t'sais, on vit dans une communauté... d'éviter la chambre à échos... mais tu perds [l'avantage] que l'algorithme te connaisse... Ça pose un conflit entre mes intérêts individuels et ce qui serait bénéfique à la collectivité.

La même participante révèle aussi, durant l'enquête, qu'elle n'a pas tendance à activer ses pouvoirs d'(in)action agentielle sur les plateformes de diffusion vidéo parce qu'en réalité,

[...] je n'ai aucune envie de prendre parti à l'espace public, puis de me prononcer... T'sais, dans la vie, je suis assez nihiliste... je [ne] vote pas. Donc, effectivement, dans mon mode d'action [sur les plateformes], c'est très similaire, je ne me prononce pas, je ne prends pas position.

Par ailleurs, une autre personne participante établit elle aussi un lien entre son « inaction » politique (elle vote, mais sans plus) et le fait qu'elle n'active pas souvent ses pouvoirs d'(in)action agentielle sur les plateformes pour y exprimer ses opinions ou son mécontentement. Comme le remarquent Derbas Thibodeau et Poirier (2019, p. 62), il ne faut en effet pas assumer que les personnes usagères souhaitent nécessairement toutes être placées en situation constante et élevée d'agentivité.

3.4 La citoyenneté culturelle

La notion de citoyenneté culturelle, qui implique la juxtaposition de principes démocratiques à l'expérience des arts et de la culture, peut aussi se révéler utile à la compréhension des motivations des personnes usagères à activer leurs pouvoirs d'(in)action agentielle dans le contexte de leur relation à la recommandation algorithmique. À cet égard, nous pouvons retenir le portrait que dresse Poirier (2017, p. 165) de la citoyenneté culturelle, qui indique que cette dernière se situe entre autres « dans le prolongement de la démocratie culturelle », une démocratie aux volets aussi bien verticaux, du fait d'une relation qui y est opérée entre les individus et les institutions, que transversaux, vu les relations d'individu à individu qu'elle habilite par sa création d'un espace dialogique. Les résultats de cette enquête illustrent en effet que les personnes participantes activent leurs pouvoirs d'(in)action agentielle aussi bien pour entrer en relation verticale avec l'institution culturelle que représente la plateforme de diffusion vidéo en continu que pour entrer dans un espace public marqué de relations transversales entre consommateurs et consommatrices de produits culturels.

À propos du rapport vertical entre l'individu consommateur de culture et l'institution culturelle, Derbas Thibodeau et Poirier (2019, p. 55-56) remarquent qu'une verticalité descendante, qui est souvent critiquée, « se double [toutefois] d'un rapport ascendant des citoyens vers l'institution, voire d'une multiplication et d'une délinéarisation des rapports », les personnes citoyennes entrant en relation, d'une multitude de manières et parfois de leur propre initiative, avec l'institution culturelle. Trois des personnes participantes remarquent justement que c'est important pour elles d'exprimer leurs jugements culturels aux plateformes, c'est-à-dire de signifier à ces plateformes, d'une manière ou d'une autre, les contenus et genres qu'elles apprécient – de leur « partager » ou de leur indiquer ce qu'elles considèrent comme des goûts culturels légitimes (Lahire, 2004). Il s'agit pour elles d'une manière explicite d'encourager les producteurs

et productrices de contenus culturels qu'elles apprécient, mais, aussi, de façon plus implicite, d'une manière pour elles d'inciter les plateformes à créer, diffuser ou rendre plus visibles des contenus qui les intéressent, *elles*, un processus décrit par Siles et ses collaborateurs (2019) comme contribuant à une « domestication mutuelle » entre la personne usagère et la plateforme de diffusion en continu, dans le cadre de laquelle les personnes usagères et les plateformes doivent s'adapter mutuellement à la forme et au « comportement » de l'autre.

Ainsi, en activant leurs pouvoirs d'action agentielle sur ces plateformes, les personnes usagères empruntent partiellement la position d'une critique culturelle, dont les goûts et les opinions sont pris en compte par les institutions culturelles. Effectivement, dans les champs culturels numériques, toutes les personnes usagères peuvent en quelque sorte se proposer comme critiques d'art, remarque Roberge (2011) – il ne s'agit pour elles que de cliquer sur un bouton d'appréciation ou d'écrire un commentaire, là où c'est rendu possible par la structure technique de l'interface d'usage, pour emprunter une telle position.

Par ailleurs, Kant, dans la *Critique du jugement*, insiste que le jugement de goût tire sa validité du fait d'un accord potentiel avec autrui plutôt que d'un simple accord avec soi-même (Arendt, 1972), et, effectivement, les résultats de l'enquête relèvent que les personnes participantes ne cherchent pas qu'à « convaincre » les plateformes de la validité ou de l'importance de leurs goûts, mais qu'elles accordent aussi de l'importance à la relation dialogique qu'elles peuvent entretenir avec d'autres personnes usagères au sein de ce qu'on pourrait nommer la « sphère esthétique publique » (Roberge, 2011 ; Wright, 2015). Arendt souligne d'ailleurs qu'une faculté associée à la culture est celle de transcender les individualités et d'instaurer des postures dialogiques (Arendt, 1972) dans lesquelles les citoyens et citoyennes de la culture « sortent » de leur perspective solipsiste (Derbas Thibodeau et Poirier, 2019) et développent la capacité de comprendre la perspective de l'« Autre » (Poirier, 2017).

C'est dans cette logique que se situe une participante, qui souligne à quel point elle apprécie pouvoir lire les commentaires d'autres personnes usagères sous les vidéos qu'elle visionne sur YouTube. À ce propos, elle explique :

Tu peux voir dans les commentaires s'il y a des gens qui ont des choses à dire par rapport à être pour ou contre... Ça te permet, justement, d'avoir une réflexion

personnelle en tant que citoyen, t'sais, c'est plus démocratique... En fait, je me rends compte... les plateformes sans commentaires, c'est presque comme des plateformes... totalitaires, j'ai l'impression, de la manière que je suis en train de voir ça [rires].

L'(in)action agentielle liée à la recommandation algorithmique sur de telles plateformes est donc perçue comme une occasion d'exercer sa citoyenneté culturelle. Ainsi, le désir d'être une citoyenne ou un citoyen d'un champ culturel, et d'activer cette citoyenneté, est un autre facteur favorisant l'(in)action agentielle des personnes usagères face à la recommandation culturelle algorithmique.

4. FACTEURS FREINANT L'(IN)ACTION AGENTIELLE LIÉE À LA RECOMMANDATION CULTURELLE ALGORITHMIQUE

4.1 La satisfaction en rapport avec la recommandation algorithmique

Un des facteurs contribuant à la décision que prennent certaines personnes usagères de ne pas activer leurs pouvoirs d'(in)action agentielle dans l'optique de « répondre » (*talk back*), de déjouer ou de manipuler les algorithmes de recommandation est la satisfaction qu'elles peuvent ressentir par rapport au fait qu'elles reçoivent des recommandations culturelles « personnalisées » et « adaptées » à leurs désirs : elles sont satisfaites du service offert par la recommandation algorithmique et ont confiance en la capacité du système automatisé à satisfaire leurs désirs de consommation (Arielli, 2018). Effectivement, même si plusieurs personnes participantes reconnaissent plusieurs limites de la recommandation algorithmique, elles y voient tout de même toutes certains avantages, alors qu'elle leur permet entre autres, dans un contexte de surcharge informationnelle, d'orienter leur attention vers des contenus susceptibles de les intéresser (Ménard, 2014). Ainsi, il s'avère que la satisfaction ou la « complaisance » (pour reprendre le terme employé par une personne participante) de certaines personnes usagères en relation avec la recommandation algorithmique sur les plateformes de diffusion vidéo en continu serait un « auxiliaire [de leur] désengagement civique » (Greffet et Wojcik, 2014, p. 145), un facteur portant entrave à leur motivation d'activer leurs pouvoirs d'(in)action agentielle par rapport au système de recommandation algorithmique.

4.2 Le coût temporel et cognitif de l'activation des pouvoirs agentiels

Un autre facteur portant entrave à la motivation des usagers et usagères d'activer leurs pouvoirs d'(in)action agentielle pour influencer les algorithmes de recommandation est le coût temporel et cognitif de ces actions.

En matière de coût temporel, les résultats de l'enquête illustrent que, même si les temporalités qui sont jaugées – celles que demandent l'activation des différents pouvoirs d'action offerts par les plateformes aux personnes usagères – sont minimales (il est souvent question d'actions requérant quelques secondes seulement), il n'en demeure pas moins qu'elles représentent un coût que certaines jugent assez considérable lors de leur analyse coût-bénéfice pour les faire pencher vers la « passivité » ou l'inactivation par défaut (non agentielle) de leurs pouvoirs d'action. En effet, une personne participante remarque : *« Quand je suis sur YouTube, c'est court, c'est rapide, je veux juste trouver quelque chose, cliquer dessus, pis le regarder. Fait que je ne vais pas faire une action [comme masquer des contenus, par exemple], parce que je veux juste consommer un truc rapidement »*.

Ce type de consommation, marqué par la « passivité » ou l'inactivation par défaut de pouvoirs d'actions, se voit donc modulé par la structure technologique qu'est l'interface de la plateforme, et implique par ailleurs que les personnes usagères passent moins de temps à affiner leurs capacités de choix et de jugement culturels, qui est pourtant, selon Lawrence (2015), ce qui procurerait aux consommateurs et consommatrices de la culture le plus de plaisir esthétique à long terme. Une consommation culturelle dite « passive », dictée par les algorithmes de recommandation, est en effet parfois empruntée par les personnes participantes, dont le choix culturel est pour ainsi dire quasiment évacué de l'expérience de consommation culturelle « grâce » à l'« épargne » temporelle que promet la recommandation algorithmique (Wright, 2015; Barth *et al.*, 2019; Bergström, 2015).

Par ailleurs, il ne s'agit pas seulement du coût temporel, mais aussi du coût cognitif que comporte l'activation des pouvoirs agentiels. À ce propos, Olive et ses collaborateurs (1997) expliquent que le concept du « coût cognitif », ou le calcul de l'« effort cognitif », réfère à la quantité de ressources attentionnelles et énergétiques qui sont exigées pour traiter une pensée ou établir une stratégie pour résoudre une tâche particulière. Effectivement, il s'avère que les personnes participantes modulent parfois leur (in)activation de leurs pouvoirs agentiels sur la base d'une estimation du coût cognitif que leur demanderaient ces actions. Une participante exprime d'ailleurs qu'elle n'active pas souvent ses pouvoirs d'(in)action agentielle

sur les plateformes de diffusion vidéo en continu parce que *« l'activité de lurking [de rodage], c'est moins forçant que d'écrire un commentaire, ou de cliquer sur like... même si ça semble ridicule »*. Un autre participant reprend le champ sémantique du « coût » ou de la « transaction » en expliquant qu'il n'effectue aucune action pour indiquer aux algorithmes qu'un contenu ne l'intéresse pas *« ...peut-être [parce] que c'est trop cher de clics »*. La raison qu'il invoque pour demeurer « passif », c'est donc *« l'engagement que ça demande, ou l'effort que ça demande »*, dit-il. Selon Ménard (2014), en choisissant de la sorte de conserver leurs ressources cognitives plutôt que de les « dépenser » pour l'activation de pouvoirs d'(in)action agentielle, ces personnes usagères délèguent leur rationalité au système, auquel elles laissent conséquemment le soin de prendre les « bons » choix esthétiques ou culturels (Arielli, 2018).

4.3 Le désir de protection de la vie privée et de l'intimité

Les résultats de l'enquête illustrent que le désir de protection de la « vie privée » ou de l'« intimité » est un autre facteur pouvant décourager les personnes usagères d'activer leurs pouvoirs d'(in)action agentielle en relation à la recommandation algorithmique sur les plateformes de diffusion vidéo en continu.

Bien que quelques-unes des personnes participantes aient démontré avoir conscience du fait que leurs données sont collectées et conservées par les organisations (souvent privées et multinationales) auxquelles appartiennent les plateformes de diffusion continue qu'elles fréquentent, leurs préoccupations en lien avec la protection de la confidentialité de leur « vie privée » n'ont pas trait aux ambitions commerciales ou manipulatrices des organisations possédant les données les concernant (Lanier, 2018 ; Seaver, 2019 ; Kalpokas, 2019), ou au potentiel de surveillance numérique auquel elles font face (Peacock, 2014), mais plutôt au jugement des individus avec lesquels elles entretiennent des relations interpersonnelles – les gens qu'elles connaissent personnellement. En effet, certaines personnes participantes limitent leurs usages et l'activation de certains de leurs pouvoirs d'actions sur ces plateformes afin d'éviter que leurs pratiques de consommation culturelle soient remarquées et jugées par leurs proches ou les individus qu'elles connaissent. Plusieurs personnes participantes ont d'ailleurs révélé être quelquefois induites, par leur désir de protection de la confidentialité de leurs pratiques de consommation culturelle, à adopter des pratiques qu'elles situent dans un paradigme d'invisibilité (comme le déroulement d'écran rapide signalant un « désintérêt » ou le clic signalant

un intérêt), plutôt que des pratiques qu'elles associent à un paradigme de visibilité, telles que la rédaction d'un commentaire ou la sélection d'un pouce pointant vers le haut.

Il s'avère donc que ce que Lahire (2004) appelle la « honte culturelle », qui diminuerait selon lui lorsque la consommation culturelle est privée plutôt que publique (regarder la télévision seul, seule ou en couple plutôt qu'aller au cinéma, par exemple), semble prendre du terrain dans des environnements numériques, où les frontières entre le privé et le public sont changeantes et ambiguës (Greffet et Wojcik, 2014 ; Azzaria et Zimmermann, 2019 ; Manach, 2010 ; Bloche et Verchère, 2011 ; Casemajor *et al.*, 2021) et que le désir de protection de la « vie privée » ou de la confidentialité des usages rend les personnes usagères peu enclines à activer certains de leurs pouvoirs d'(in)action agentielle face à la recommandation algorithmique.

4.4 Du défaitisme quant à la portée des actions

Finalement, un dernier facteur freinant l'(in)action agentielle liée à la recommandation algorithmique de contenus culturels est le défaitisme, provoqué par un sentiment d'impuissance, que certaines personnes usagères ressentent face aux systèmes algorithmiques, sentiment qui porte vraisemblablement entrave à leur motivation d'activer leurs pouvoirs d'(in)action agentielle sur les plateformes pour influencer, interagir, « domestiquer » ou « répondre » aux algorithmes de recommandation.

Effectivement, une participante souligne que ses pouvoirs agentiels lui semblent minimes et dotés de peu de potentiel d'influence. Elle explique qu'un facteur portant entrave à son désir d'activer ces pouvoirs est

le fait d'être petit dans ce grand monde... L'immensité de la chose... Ma petite personne face à l'ensemble, je pense que ça c'est [un facteur]. Puis aussi, l'immensité de tout ce qui est offert... la quantité du contenu, la quantité est vraiment présente.

Certaines personnes usagères ressentent en effet qu'en n'étant qu'un minime agrégat de données dans l'espace social et symbolique qu'est le système algorithmique, qui est hôte d'une quantité massive de données, elles ne peuvent pas réellement y être prises en compte en tant qu'individus complexes et multidimensionnels, et ont conséquemment l'impression que leurs actions individuelles n'ont pas de portée sur le système algorithmique, auquel elles aimeraient toutefois pouvoir rétorquer ou répondre afin « d'ouvrir le carcan » que représente pour elles la recommandation algorithmique culturelle et sa modalité filtrante.

Ces personnes participantes adoptent une posture défaitiste, dans laquelle la portée effective et symbolique que pourrait représenter l'activation de leurs pouvoirs d'(in)action agentielle est minimisée, ce qui, évidemment, les mène à moins se prévaloir desdits pouvoirs. Hoffman et ses collaborateurs (2016) réfèrent justement au fait que plusieurs personnes usagères des plateformes numériques, aux prises avec des forces qu'elles considèrent écrasantes, ressentent qu'elles ne disposent d'aucun contrôle sur les plateformes qu'elles fréquentent et sur les systèmes auxquels elles participent, ce qui cause chez elles un sentiment d'impuissance, souvent doublé d'une méfiance envers les motivations et les intérêts des plateformes. Milan et Tréré (2019) mentionnent pour leur part un sentiment d'inéluctabilité envers la donnéification, sentiment qui serait en voie d'être généralisé et qui empêcherait l'imagination d'alternatives possibles à ce phénomène technosocial.

Une telle vision est corroborée par des résultats de cette enquête, qui illustrent qu'un sentiment d'impuissance mène certaines personnes participantes à être défaitistes et, ainsi, à tendre vers l'inactivation de leurs pouvoirs d'(in)action agentielle, même si elles se révèlent être convaincues que la recommandation algorithmique peut être « *effrayante* », « *vicieuse* » ou « *dangereuse* », pour reprendre les termes que certaines utilisaient vers la fin de leur deuxième entretien. Une personne participante, qui a adopté une position critique face aux algorithmes de recommandation au cours de l'enquête, a justement affirmé qu'elle demeurera tout de même « *passive* », qu'elle continuera à peu ou ne pas activer ses pouvoirs agentiels pour influencer le système algorithmique, car elle est « *un peu cynique* » et a peu confiance en l'habileté des personnes consommatrices ou citoyennes à être porteuses de changement. Une autre se dit un peu « *nihiliste* », et dit, comme pour justifier l'inaction qu'elle maintient face à la structure algorithmique dans laquelle s'inscrit sa consommation culturelle : « *On est dedans... T'sais, on vit là-dedans.* »

CONCLUSION

Ce chapitre a abordé la question des dilemmes agentiels des individus usagers de plateformes de diffusion vidéo en continu face au choix d'activer, ou non, les possibilités d'(in)actions agentielles liées à la recommandation algorithmique qui leur sont offertes par les plateformes. La proposition classificatoire de ces possibilités d'(in)actions agentielles se voit bien entendu contrainte par les limites des démarches méthodologique

et analytique empruntées : notamment, par un biais de sélection dans la constitution de l'échantillon de personnes participantes induit par la sélection d'individus positionnés dans le cercle de connaissances de la chercheuse ; par le nombre limité de personnes participantes ; par un biais possible de désirabilité sociale dans certaines réponses de ces dernières ; par une construction de données et de discours n'étant pas préexistants, chez les personnes participantes, à la réalisation de l'étude, étant donné l'approche de coconstruction de sens empruntée et l'émancipation du pouvoir réflexif des personnes participantes au cours des entretiens ; et par l'approche subjective, socialement et épistémologiquement située, de la chercheuse dans sa manière de construire l'étude, de conduire les entretiens et d'en analyser les résultats. Toutefois, il est proposé que la reconnaissance explicite de ces limites permet, dans une certaine mesure, d'établir les contours de la validité empirique et qualitative des résultats obtenus (Selz et Maillochon, 2009).

En somme, les personnes usagères peuvent développer des réflexions et des positions critiques, et peuvent activer leurs pouvoirs d'action face à la recommandation algorithmique sur les plateformes de diffusion vidéo en continu, mais il n'en demeure pas moins qu'elles sont prises *dans* ces structures technosociales (Seaver, 2019) – la condition même de leur agentivité au sein des systèmes de recommandation algorithmique est effectivement leur présence au sein de ceux-ci, qui ont donc la capacité structurante d'établir les limites de l'agentivité possible en leur sein (Giddens, 1984 ; Layder, 2006). Ainsi, les personnes usagères se voient aux prises avec des dilemmes agentiels constants en relation avec les systèmes de recommandation algorithmique, au sein desquels leur posture agentielle est non seulement tiraillée entre différentes impulsions d'(in)action issues des facteurs proposés ci-dessus, mais est aussi confrontée à l'architecture technique des plateformes, qui régit le domaine des (in)actions qui y sont possibles.

Compte tenu de la technologie d'apprentissage profond régissant plusieurs systèmes algorithmiques, technologie qui fonctionne d'une telle manière que les données y entrant (l'*input*) modifient continuellement les calculs et les résultats de ces systèmes, il est toutefois pertinent de soulever que les comportements numériques des personnes usagères – les données d'usage qu'elles produisent – structurent réciproquement les plateformes et leurs données de sortie (*outputs*) algorithmiques. Cette dynamique, partie constituante d'une relation dite de « domestication mutuelle » (Siles *et al.*, 2019) ou « mutuellement constitutive » (McKelvey et Hunt, 2019) entre les personnes usagères et les plateformes faisant usage de la

recommandation algorithmique en apprentissage profond, implique ainsi que la structuration de l'agir serait partiellement réciproque, avec, d'une part, les plateformes contraignant les possibilités d'action des personnes usagères et encourageant fortement certains comportements chez elles par l'incitation comportementale (*nudging*) algorithmique (Kalpokas, 2019), et, d'autre part, les personnes usagères contribuant à donner forme aux manifestations des systèmes algorithmiques en les « nourrissant » de leurs données d'usage et en influençant ainsi leurs possibilités de développement formel. Il serait ainsi pertinent d'étudier la possibilité que dans de tels assemblages, au sein desquels la structuration et l'agir se font miroitants, l'agentivité, plus encore que contrainte, est potentiellement fragmentée et distribuée.

RÉFÉRENCES

- Arendt, Hannah, *La crise de la culture*, Paris, Gallimard-Folio, 1972.
- Arielli, Emanuele, « Taste and the algorithm », *Studi di estetica*, vol. XLVI, série IV, n° 3, 2018, p. 77-97.
- Azzaria, Georges et Hélène Zimmermann, « Pratiques et représentations des internautes québécois en matière de consentement, de droit à la vie privée et de droit d'auteur », *Lex Electronica*, n° 24, Hors-série, 2019.
- Baringhorst, Sigrid et Mundo Yang, « Types of creative citizenship : The engagement of consumer netizens », *Symposium ECREA : (Mis)understanding Political Participation*, Munich, Université LM, 2013.
- Barth, Susanne, Menno D.T. de Jong, Marianne Junger, Pietier Hartel et Janina Roppelt, « Putting the privacy paradox to the test : Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources », *Telematics and Informatics*, n° 41, 2019, p. 55-69.
- Beer, David, *Popular Culture and New Media : The Politics of Circulation*, Londres, Palgrave-Macmillan, 2013.
- Bellavance, Guy, « Where's high ? Who's low ? What's new ? Classification and stratification inside cultural "repertoires" », *Poetics (Journal of Empirical Research on Culture, Media and the Arts)*, vol. 36, n°s 2-3, 2008, p. 189-216.
- Bergström, Annika, « Online privacy concerns : A broad approach to understanding the concerns of different groups for different uses », *Computers in Human Behavior*, n° 53, 2015, p. 419-426.
- Beuscart, Jean-Samuel, Samuel Coavoux et Sisley Maillard, « Les algorithmes de recommandation musicale et l'autonomie de l'auditeur. Analyse des écoutes d'un panel d'utilisateurs de streaming », *La Découverte*, n° 213, 2019, p. 17-47.
- Bloche, Patrick et Patrice Verchère, *Rapport d'information sur les droits de l'individu dans la révolution numérique*, Assemblée nationale française, Document n° 3560, 2011.

- Casemajor, Nathalie, Guillaume Sirois, Lily-Cannelle Mathieu, Alexis Morin et Louis Mélançon, *La gouvernance des données d'usage : enjeux éthiques et perceptions des publics dans les bibliothèques et archives. Synthèse de la littérature*, Montréal, Institut national de la recherche scientifique, 2021.
- Claes, Arnaud, Victor Wiard, Heidi Mercenier, Thibault Philippette, Marie Dufrasne, Arnaud Browet et Raphael Jungers, « Algorithmes de recommandation et culture technique : penser le dialogue entre éducation et design », *tic&société*, vol. 15, n° 1, 2021, p. 127-157.
- Couldry, Nick, « Inaugural : A necessary disenchantment : Myth, agency and injustice in a digital world », *The Sociological Review*, vol. 62, n° 4, 2014, p. 880-897.
- Coûtant, Alexandre et Jean-Claude Domenget, « L'analyse du numérique dans une perspective complexe avec la sociologie des usages », dans Mélanie Millette, Florence Millerand, David Myles et Guillaume Latzko-Toth (dir.), *Méthodes de recherche en contexte numérique : une orientation qualitative*, Montréal, Les Presses de l'Université de Montréal, 2020.
- Derbas Thibodeau, François R. et Christian Poirier, « Bibliothèques publiques et virage citoyen : enjeux institutionnels et communicationnels », *Communiquer, Revue de communication sociale et publique*, n° 26, 2019, p. 47-66.
- Finn, Ed, *What Algorithms Want : Imagination in the Age of Computing*, Cambridge, MIT Press, 2017.
- Frischmann, Brett, Contribution dans le panel « Rethinking the Internet : How we lost control and how to take it back », *World Science Festival*, 2018, <https://www.youtube.com/watch?v=sIsX_wihgCY>.
- Gallant, Nicole, Katherine Labrecque, Guillaume Latzko-Toth et Madeleine Pastinelli, « La visite commentée : documenter les pratiques numériques par l'entretien sur traces », dans Mélanie Millette, Florence Millerand, David Myles et Guillaume Latzko-Toth (dir.), *Méthodes de recherche en contexte numérique : une orientation qualitative*, Montréal, Les Presses de l'Université de Montréal, 2020.
- Gallant, Nicole, Guillaume Latzko-Toth et Madeleine Pastinelli, *Circulation de l'information sur les médias sociaux pendant la grève étudiante de 2012 au Québec*, Montréal, Institut national de la recherche scientifique, Centre d'études des médias et CMRC-CCRM, 2015.
- Giddens, Anthony, *The Constitution of Society : Outline of the Theory of Structuration*, Berkeley, University of California Press, 1984.
- Gillespie, Tarleton, « The relevance of algorithms », dans Tarleton Gillespie, Pablo J. Boczkowski et Kirsten A. Foot (dir.), *Media Technologies : Essays on Communication, Materiality and Society*, Cambridge, MIT Press, 2014.
- Gillespie, Tarleton, « #trendingistrending. When algorithms become culture », dans Robert Seyfert et Jonathan Roberge (dir.), *Algorithmic Cultures : Essays on Meaning, Performance and New Technologies*, Milton Park, Routledge, 2016.
- Greffet, Fabienne et Stéphanie Wojcik, « La citoyenneté numérique : perspectives de recherche », *Réseaux*, vol. 2, n°s 184-185, 2014, p. 125-159.
- Gubrium, Jaber F. et James A. Holstein, « Narrative practice and the transformation of interview subjectivity », dans Jaber F. Gubrium, James A. Holstein, Amir B. Marvasti et Karyn D. McKinney (dir.), *The SAGE Handbook of Interview Research : The Complexity of the Craft*, Thousand Oaks, SAGE Publications Ltd., 2012, p. 27-43.
- Hallinan, Blake et Ted Striphas, « Recommended for you : The Netflix Prize and the production of algorithmic culture », *New Media & Society*, vol. 18, n° 1, 2016, p. 117-137.

- Hoffman, Christian Pieter, Christoph Lutz et Giulia Ranzini, « Privacy cynicism : A new approach to the privacy paradox », *Journal of Psychosocial Research on Cyberspace*, vol. 10, n° 4, article 7, 2016.
- Hutchison, Jonathan, *Cultural Intermediaries : Audience Participation in Media Organisations*, Londres, Palgrave Macmillan, 2017.
- Introna, Lucas Daniel, « The algorithmic choreography of the impressionable subject », dans Robert Seyfert, et Jonathan Roberge (dir.), *Algorithmic Cultures : Essays on Meaning, Performance and New Technologies*, Milton Park, Routledge, 2016.
- Jouët, Josiane et Coralie Le Caroff, « L'observation ethnographique en ligne », dans Christine Barats (dir.), *Manuel d'analyse du web en sciences humaines et sociales*, 2^e édition actualisée et complétée, Paris, Armand Colin, 2016, p. 156-175.
- Kalpokas, Ignas, *Algorithmic Governance*, Londres, Palgrave Pivot, 2019.
- Karakayali, Nedim, Burc Kostem et Idil Galip, « Recommendation systems as technologies of the self : algorithmic control and the formation of music taste », *Theory, Culture & Society*, vol. 35, n° 2, 2018, p. 3-24.
- Karpik, Lucien, *L'économie des biens singuliers*, Paris, Gallimard, 2007.
- Kennedy, Helen, Thomas Poell et Jose van Dijck, « Data and agency », *Big Data & Society*, juillet-décembre 2015, p. 1-7.
- Kitchin, Rob et Martin Dodge, *Code/Space : Software and Everyday Life*, Cambridge, MIT Press, 2011.
- Kitchin, Rob et Tracey Lauriault, « Towards critical data studies : Charting and unpacking data assemblages and their work », *The Programmable City*, Working Paper 2, 2014, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112>.
- Lahire, Bernard, *La culture des individus : dissonances culturelles et distinctions de soi*, Paris, Éditions La Découverte, 2004.
- Lamoureux, Diane, « Hannah Arendt, l'esthétique et le politique », *Revue québécoise de science politique*, n° 25, 1994, p. 65-87.
- Lanier, Jaron, Contribution dans le panel « Rethinking the Internet : How we lost control and how to take it back », *World Science Festival*, 2018, <https://www.youtube.com/watch?v=sIsX_wihgCY>.
- Latzko-Toth, Guillaume, Claudine Bonneau et Mélanie Millette, « Small data, thick data : thickening strategies for trace-based social media research », dans Luke Sloan et Anabel Quan-Haase (dir.), *The SAGE Handbook of Social Media Research Methods*, Thousand Oaks, SAGE Publications Inc., 2017.
- Lawrence, Emily, « Everything is a recommendation. Netflix, altgenres and the construction of taste », *Knowledge Organization*, vol. 42, n° 5, 2015, p. 358-364.
- Layder, Derek, *Understanding Social Theory*, 2^e édition, Londres, SAGE Publications Ltd., 2006.
- Logg, Jennifer, Julia Minson et Don A. Moore, « Algorithm appreciation : People prefer algorithmic to human judgment », *Organizational Behavior and Human Decision Processes*, n° 151, 2019, p. 90-103.
- Manach, Jean-Marc, *La vie privée, un problème de vieux cons ?*, Limoges, Fyp Éditions, 2020.
- McKelvey, Fenwick et Robert Hunt, « Discoverability : Toward a definition of content discovery through platforms », *Social Media + Society*, janvier-mars 2019, p. 1-15.
- Ménard, Marc, « Systèmes de recommandation de biens culturels : vers une production de conformité? », *Les Cahiers du Numérique*, vol. 1, n° 10, 2014, p. 69-94.

- Milan, Stefania et Emiliano Treré, « Big data from the South(s) : Beyond data universalism », *Television & New Media*, vol. 20, n° 4, 2019, p. 319-355.
- Mota Drumond et Gabrielle Silva, « La boîte noire de Netflix : la configuration des usages par les interfaces et les algorithmes », dans Mélanie Millette, Florence Millerand, David Myles et Guillaume Latzko-Toth (dir.), *Méthodes de recherche en contexte numérique : une orientation qualitative*, Montréal, Les Presses de l'Université de Montréal, 2020.
- Olive, Thierry, Annie Piolat et Jean-Yves Roussey, *Effort cognitif et mobilisation des processus en production de texte : effet de l'habileté rédactionnelle et du niveau des connaissances*, Université de Provence, Centre de recherche PsyCLE – EA 3273, 1997, <<http://centrepysycle-amu.fr/wp-content/uploads/2014/01/OlivePR1997.pdf>>.
- Organisation de coopération et de développement économiques (OCDE), *La littératie à l'ère de l'information : rapport final de l'enquête internationale sur la littératie des adultes*, Paris, OCDE et Statistique Canada, 2000, <<http://www.oecd.org/fr/education/innovation-education/39438013.pdf>>.
- Ovadya, Aviv, Contribution dans le panel « Rethinking the Internet : How we lost control and how to take it back », *World Science Festival*, 2018, <https://www.youtube.com/watch?v=sIsX_wihgCY>.
- Passeron, Jean-Claude, « Acteur, agent, actant : personnages en quête d'un scénario introuvable », *Revue européenne des sciences sociales*, vol. XXXIX, n° 121, 2001, p. 15-30.
- Peacock, Sylvia E., « How web tracking changes user agency in the age of big data : The used user », *Big Data & Society*, vol. 1, n° 2, 2014, p. 1-11.
- Pedrycz, Witold et Shyi-Ming Chen (dir.), *Social Networks : A Framework of Computational Intelligence*, New York, Springer, 2013.
- Poirier, Christian, « La citoyenneté culturelle. Considérations théoriques et empiriques », dans Nathalie Casemajor, Marcelle Dubé, Jean-Marie Lafortune et Ève Lamoureux (dir.), *Expériences critiques de la médiation culturelle*, Québec, Presses de l'Université Laval, 2017.
- Roberge, Jonathan, « The aesthetic public sphere and the transformation of criticism », *Social Semiotics*, vol. 21, n° 3, 2011, p. 435-453.
- Rollet, Nicolas, Valérie Beaudouin et Isabelle Garron, « Vidéo-ethnographie des usages de Gallica : une exploration au plus près de l'activité », *Document numérique*, vol. 2-3, n° 20, 2017, p. 97-114.
- Rosanvallon, Pierre, *La contre-démocratie : la politique à l'âge de la défiance*, Paris, Seuil, 2006.
- Savoie-Zajc, Lorraine, « L'entrevue semi-dirigée », dans Benoît Gauthier et Isabelle Bourgeois (dir.), *Recherche sociale : de la problématique à la collecte de données*, 6^e édition, Québec, Presses de l'Université du Québec, 2016.
- Schudson, Michael, *The Good Citizen : A History of American Civic Life*, New York, Free Press, 1998.
- Seaver, Nick, « Algorithms as culture : Some tactics for the ethnography of algorithmic systems », *Big Data & Society*, juillet-décembre 2017, p. 1-12.
- Seaver, Nick, « Captivating algorithms : Recommender systems as traps », *Journal of Material Culture*, vol. 24, n° 4, 2019, p. 421-436.
- Selz, Marion et Florence Maillachon, *Le raisonnement statistique en sociologie*, Paris, Presses universitaires de France, 2009.

- Siles, Ignacio, Johan Espinoza-Rojas, Adrian Naranjo et Maria Fernanda Tristan, « The mutual domestication of users and algorithmic recommendations on Netflix », *Communication, Culture & Critique*, n° 12, 2019, p. 499-518.
- Small, Mario Luis, « How many cases do I need? On science and the logic of case selection in field-based research », *Ethnography*, vol. 10, n° 1, 2009, p. 5-38.
- Wright, David, « Digitalizing tastes », dans *Understanding Cultural Taste: Sensation, Skill and Sensibility*, Londres, Palgrave Macmillan, 2015.
- Ylipulli, Johanna et Aale Luusua, « Without libraries what have we? : Public libraries as nodes for technological empowerment in the era of smart cities, AI and big data », *Proceedings of the 9th International Conference on Communities & Technologies*, 2019, p. 92-101.

QUAND LES ABONNÉS ET ABONNÉES TRAVAILLENT LEUR PROFIL

NETFLIX, ALGORITHMES ET PROMESSES DE DÉCOUVRABILITÉ

**Alix Bénistant¹, Christine Chevret-Castellani²
et Sarah Labelle³**

Les pratiques de consommation culturelle comme l'écoute musicale ou le visionnage audiovisuel se distinguent les unes des autres par leur rapport à la découverte de produits. Dans le domaine musical, l'écoute répétée de morceaux tout comme l'élaboration et le partage de listes d'écoute constituent le « principe moteur de la dynamique du goût musical et de l'attachement à la musique » (Maisonneuve, 2019) et les pratiques « où la réalisation d'expériences réussies de découverte est un enjeu central pour l'amateur » (*ibid.*). Dans le domaine audiovisuel s'observe une distinction forte : si la cinéphilie sous-entend la pratique de voir à plusieurs reprises films ou séries, la consommation audiovisuelle se caractérise davantage par le passage d'une production à une autre. L'idée de passer à la production suivante est pleinement intégrée dans les logiques de présentation des catalogues et les logiques de recommandation. Le slogan de Netflix n'est-il

-
1. Maître de conférences en sciences de l'information et de la communication à l'Université Sorbonne Paris Nord.
 2. Enseignante-chercheuse en sciences de l'information et de la communication à l'Université Sorbonne Paris Nord.
 3. Professeure des Universités en sciences de l'information et de la communication à l'Université Paul Valéry Montpellier 3.

pas « *See What's Next* » ? S'observe un ici et maintenant de la consommation qui suppose que la recommandation joue un rôle dans le futur immédiat de la pratique de consommation et qui repose sur la capacité du dispositif à sans arrêt suggérer de nouvelles productions.

Dans cet article, nous avons fait le choix d'analyser les pratiques de jeunes consommateurs et les stratégies algorithmiques mises en place par une plateforme de distribution de contenus en flux continu d'un acteur spécifique de l'audiovisuel : Netflix. Certes, porter la focale sur un acteur majeur de la filière bénéficiant déjà d'une visibilité importante liée à une position dominante sur le marché mondial⁴ a l'inconvénient de renforcer l'attention sur celui-ci. Mais cette démarche présente l'avantage, lorsqu'on souhaite étudier tant le dispositif que les pratiques de consommation le concernant, d'observer un acteur bien implanté et de pouvoir dès lors enquêter auprès d'abonnés de sa plateforme ayant une expérience et une pratique relativement approfondie de celle-ci.

Par ailleurs, s'il est certain que les outils de recommandation sont loin d'être les seuls mobilisés par les usagers dans la panoplie dont ils disposent (panoplie qui se déploie largement en dehors du dispositif de la plateforme⁵), il ne nous semble pas moins primordial de les examiner afin de comprendre leur rôle – jusqu'ici exploré davantage dans une démarche de rétro-ingénierie sociotechnique que du côté des usagers – dans ce que les recherches à leur propos nomment « économie de la jouissance » (Coutant *et al.*, 2018) ou « économie de la découverte » (Maisonnette, 2019). Nous chercherons donc à voir dans quelle mesure la découvrabilité, notion fréquemment employée par les acteurs socioéconomiques, est travaillée à différents niveaux dans les logiques de recommandation portées par les dispositifs d'intelligence artificielle, en particulier les médiations algorithmiques.

Les logiques de recommandation constituent une part appréciable des conditions de la consommation culturelle, car elles reposent non seulement sur le besoin des consommateurs d'être guidés dans une offre pléthorique,

4. Netflix revendique être présent dans 190 pays. Netflix, « About Netflix », <<https://about.netflix.com/fr>> (consulté le 13 janvier 2021).

5. Sophie Maisonnette (2019), dans une étude sur la découverte musicale, montre que celle-ci se base sur différents dispositifs qu'elle présente en quatre grandes catégories (dispositifs de recommandation explicites, dispositifs de cadrage de l'offre, dispositifs de confiance et réseaux de confiance interpersonnels), en pointant que les liens sociaux « forts » (reprenant ici Granovetter) restent centraux dans les recommandations que suivent les usagers, là où les algorithmes ne rempliraient « qu'imparfaitement » cette fonction de recommandation.

mais encore sur la revendication de Netflix de savoir guider ses abonnés et abonnées dans cette offre. Partant de ce constat largement partagé, le problème général de ce chapitre pourrait être formulé de la sorte : quelle part les traitements automatisés laissent-ils à l'exploration et à la découvrabilité des productions culturelles par les abonnés et abonnées ? Nous nous interrogerons ainsi sur la manière dont les dispositifs d'intelligence artificielle créent un environnement éditorialisé qui prend appui à la fois sur la saturation d'informations et sur la création d'interstices. D'un côté, la saturation conduit à maintenir les abonnés et abonnées dans le dispositif. De l'autre, les interstices offrent la possibilité du décalage du regard et de la découverte de productions autres que celles habituellement consommées.

Nous faisons une première hypothèse sur la tension entre, d'un côté, la valeur symbolique accordée à des contenus par les consommateurs (notoriété, goûts...) et, d'un autre côté, leur présence dans un environnement de consommation entendu comme modalité d'accès à la culture (Bouquillion *et al.*, 2013, p. 42), où se construit une nouvelle manifestation de ce qu'Hesmondhalgh et Lotz (2020) nomment le *media circulation power*⁶ avec les « promesses » que celle-ci suppose (première partie). Nous travaillerons ainsi à comprendre la manière dont la part symbolique des productions est mise en avant dans des logiques éditoriales et des médiations algorithmiques participant au renforcement de leur potentiel de consommation. Les choix en matière de présentation des catalogues, s'ils semblent relever de pratiques éditoriales, n'en dépendent pas moins de logiques computationnelles que nous considérons comme participant de « l'industrialisation des biens symboliques » (Bouquillion *et al.*, 2013).

Nous proposons une deuxième hypothèse selon laquelle les abonnés et abonnées de Netflix ont une conscience aigüe de la présence et du rôle des algorithmes dans leur consommation. Nous chercherons, à partir d'observations d'une population de jeunes étudiants et étudiantes français abonnés, à comprendre le sens conféré aux traces des algorithmes sur les écrans, c'est-à-dire aux signes identifiés par les consommateurs comme marques de la présence algorithmique et à analyser la manière dont s'établit la confiance sur le profilage et les recommandations. Nous analyserons ainsi ce que nous proposons d'appeler le travail du consommateur (Dujarier,

6. Pour les auteurs, « il s'agit essentiellement du pouvoir [exercé par les entreprises des médias impliquées dans la production et la circulation de textes audiovisuels, *via* notamment les interfaces d'écran vidéo] de diriger les publics vers certains types d'expériences et de contenus, et donc de les éloigner des autres » (Hesmondhalgh et Lotz, 2020, notre traduction).

2008) de réinterprétation des filtres algorithmiques et de production de données de consommation en lien avec un bricolage (De Certeau, 1980) dans la pratique de consommation culturelle (deuxième partie).

Par ailleurs, nous montrerons que les logiques computationnelles, mobilisées pour renforcer la valeur symbolique autour des catalogues, participent à la hiérarchisation et à la dénomination des rubriques et au changement quotidien de la « une » des comptes individuels tout en élaborant un univers symbolique qui fait écho à celui de l'abonné. Là se situe l'enjeu lié à la découvrabilité des productions : il est nécessaire d'accroître la valeur symbolique des biens culturels qui sont présents pour faire en sorte que le processus de consommation aille jusqu'au bout et soit continu. En ce sens, la découvrabilité est ici entendue, d'une part, comme l'articulation entre la pratique industrielle de référencement (production de données et de métadonnées), réalisée en amont de la diffusion d'une production pour favoriser sa découverte par les consommateurs et, d'autre part, comme la capacité des consommateurs à s'appuyer sur ces traces de référencement pour consommer des productions. Nous distinguerons la découvrabilité, en tant qu'optimisation stratégique industrielle, de la découverte de nouveaux contenus par les consommateurs, en tant que pratique d'exploration tendant à diversifier la nature des productions consommées (troisième partie).

Pour parvenir à cela, nous avons mené une enquête comprenant quinze entretiens semi-directifs approfondis et « équipés » (Souchier *et al.*, 2003 ; Maisonneuve, 2019), réalisés auprès d'abonnés et abonnées pris au sein d'une population étudiante française âgée de 20 à 25 ans⁷. Nous entendons par « équipés », dans le sillage de la méthodologie mise en place par Sophie Maisonneuve lors de son étude sur « l'économie de la découverte musicale à l'ère numérique », le fait que les répondants étaient accompagnés de leur équipement habituel de visionnage et ont ainsi pu montrer et commenter « les opérations habituelles associées à leur consommation » audiovisuelle : support utilisé, « parcours de découverte et d'appropriation » et « modalités de conservation » (Maisonneuve, 2019)⁸. De cette manière, nous amenons les enquêtés et enquêtées à la « production de

7. Cette enquête a été réalisée dans le cadre du projet de recherche Data Culture financé par le Labex ICCA (Industries culturelles et création artistique).

8. Notre pratique d'enquête s'inscrit en héritage de la recherche « Écrans et réseaux : vers une transformation des rapports à l'écrit » menée pour la Bibliothèque Publique d'Informations sur la période 2000-2002 (Souchier *et al.*, 2003).

discours sur leurs propres pratiques», permettant de garder une relation dynamique entre les deux types de données empiriques ainsi produites (Catoir-Brisson et Jankeviciute, 2014). En parallèle, nous nous sommes appuyés sur un corpus de documents hétérogènes (articles, communiqués de presse, rapports, interventions publiques) qui portent sur les algorithmes de Netflix et leur «prétention médiationnelle» (Jeanneret, 2014) dans la découverte de produits audiovisuels.

1. PROMESSES D'EXPLORATION DANS UNE OFFRE CULTURELLE

Netflix met en avant son modèle de recommandation issu du traitement algorithmique des données entrantes suivant l'argument marketing de la personnalisation grâce à la puissance de calcul. Il est ce par quoi l'entreprise s'est fait reconnaître: la personnalisation fine des contenus qui correspondrait à l'horizon d'attente des usagers (Paquien-séguy, 2017, p. 22) ici favorisée par la performance du traitement informatisé des données par rapport aux autres services de vidéos à la demande (Bensamoun et Farchy, 2020, p. 17). Parmi les traitements informatisés, trois types de recommandation basés sur la nature des données sont aujourd'hui très largement recensés et étudiés par la littérature scientifique et les acteurs publics. Joëlle Farchy et Juliette Denis (2020), s'appuyant sur un document de travail de la Direction générale des médias et des industries culturelles intitulé *Les algorithmes dans la culture*, les exposent de la sorte pour Netflix: le filtrage collaboratif, le filtrage sur le contenu et les formes hybrides. Le premier s'appuie sur l'analyse des comportements des utilisateurs aux goûts proches, et établit des «proximités statistiques» (Beuscart *et al.*, 2019) conduisant à une recommandation du type «les gens qui ont aimé ce film que vous avez regardé ont aussi aimé celui-ci» à partir des comparaisons du profil du consommateur avec celui d'autres utilisateurs ayant eu accès au même produit culturel (Farchy et Denis, 2020, p. 162). Le deuxième type de recommandation s'appuie sur les caractéristiques d'un produit qui, disposant de plus d'une centaine de repères (*tags*) attribués en interne au moyen d'une intense activité humaine de visionnage, propose une catégorisation, grâce aux métadonnées, qui peut dès lors être croisée avec les données produites par la pratique de visionnage des utilisateurs. Le troisième type propose des formes hybrides combinant les différentes modalités afin d'améliorer la performance de la recommandation, où l'un viendrait pallier les inconvénients de l'autre (le filtrage par le contenu permettrait en ce sens de favoriser la découverte d'œuvres

moins plébiscitées là où le filtrage collaboratif favoriserait par essence les productions les plus populaires).

Tout d'abord, l'argument marketing de la personnalisation grâce à la puissance de l'algorithme, mis en avant par Netflix dans sa communication afin de mieux se distinguer de la concurrence, engage l'entreprise à une première « promesse » (Auray et Vétel, 2013, p. 6) qui correspondrait à l'horizon d'attente de l'utilisateur, c'est-à-dire la correspondance entre ce qui est proposé éditorialement et ce qui est consommé. C'est le cas par exemple lorsque l'entreprise annonce que « 75 % des contenus visionnés par ses clients viennent d'une recommandation personnalisée » (Tryon, 2015 cité dans Farchy et Denis, 2020, p. 164). Pour Marie-Laure Daridan, de la Direction des affaires publiques de Netflix France, la recommandation a pour fonction d'offrir à l'utilisateur « la meilleure expérience possible⁹ » et celle-ci se réalise grâce à « l'historique de visionnage¹⁰ ». Ainsi, le discours de l'entreprise défend l'idée que c'est l'algorithme qui est susceptible de s'adapter à l'utilisateur et non le consommateur qui serait influencé, déterminé, par les recommandations générées par le dispositif.

La personnalisation, grâce à la puissance de l'algorithme revendiquée dans le cadre de la recommandation, aurait aussi pour finalité de donner à l'utilisateur la possibilité d'être guidé face à une offre culturelle pléthorique. Le dispositif est présenté comme permettant la facilitation de la consommation culturelle. Cette deuxième promesse, celle d'orienter le consommateur dans le catalogue, représente un enjeu en fonction de la manière dont l'utilisateur est déterminé : comme se situant dans une continuité par rapport à sa consommation précédente ou comme étant dans une logique de découverte. La page d'accueil de l'abonné ou l'abonnée est alors modifiée soit par l'affichage de ce dont il est susceptible de poursuivre la lecture, un film ou une série dont le visionnage a déjà commencé précédemment (« Continuer à regarder »), soit par la présentation de contenus inédits¹¹. Cette promesse n'exclut cependant plus désormais pour Netflix de réintroduire du contenu éditorialisé. C'est le cas lorsque l'entreprise défend dans sa communication le travail d'éditorialisation effectué dans son modèle de recommandation, parallèlement au traitement automatisé,

9. Propos collectés lors de l'observation de l'événement « La réforme de l'audiovisuel en France : entre nouveaux acteurs et vieilles contraintes » (Université Paris 1, 3 mars 2020).

10. *Ibid.*

11. ZDNet, « Comment Netflix compose ses algorithmes de recommandation », <<https://www.zdnet.fr/actualites/comment-netflix-compose-ses-algorithmes-de-recommandation-39851718.html>> (consulté le 9 août 2021).

et notamment son « Top 10 ». Selon la représentante de la Direction des affaires publiques de Netflix précédemment citée, l'éditorialisation est de plus en plus adoptée, avec par exemple la présentation de films de « Cinquante réalisatrices » pour la Journée internationale des droits des femmes du 8 mars. La personnalisation des contenus et l'éditorialisation restent toutefois assez restreintes sur certains aspects, eu égard à la présentation de catégories peu individualisées, comme celle des « Ajouts récents » (Delaporte, 2018). L'offre de contenus du service de vidéos à la demande (SVOD) étant relativement limitée (Dessinges et Perticoz, 2019, p. 21), tout se joue autour des « régimes de valeur » (Hallinan et Strichas, 2016, p. 120) relatifs à la culture.

La troisième promesse adressée aux consommateurs de Netflix repose en ce sens sur la promotion d'une valeur, celle qui est relative à la culture, qui passe par une réassurance du consommateur à la fois sur l'usage des données et sur une supposée qualité culturelle des contenus produits et présentés dans le catalogue. Pour Marie-Laure Daridan, les seules données personnelles collectées sont les adresses de courriel et les coordonnées bancaires. Son discours vise à dissocier la collecte des données liées à la consommation de la présentation éditorialisée des produits. Les données entrantes ne seraient donc pas « personnelles » et la présentation des produits résulterait aussi d'une « demande¹² ». En particulier, la proposition d'œuvres européennes dans le catalogue serait aujourd'hui liée à une demande du public européen. Enfin, il s'agit d'indiquer au consommateur, selon Netflix, que les données récoltées sont non seulement suffisamment « neutres » pour que celui-ci garde la mainmise sur le dispositif, mais aussi assez indicatives pour garantir la qualité culturelle des œuvres présentées dans le catalogue.

Pour Netflix, les données fournissent en effet des informations permettant la recommandation sans déterminer la création. Au-delà de la promesse à destination des usagers, ce choix de l'entreprise la positionne d'ailleurs auprès de certains acteurs de l'internet comme favorisant aujourd'hui l'accès à la connaissance et la culture. Par exemple, pour Guillaume Chaslot, expert en intelligence artificielle, si certains types de contenus comme ceux exposant des théories complotistes sont plébiscités par le public, Netflix, contrairement à YouTube, ne s'engagerait pas pour

12. Propos collectés lors de l'observation de l'événement « La réforme de l'audiovisuel en France : entre nouveaux acteurs et vieilles contraintes » (*op. cit.*).

autant dans la création de documentaires qui pourraient conduire à nourrir ces théories¹³. Pour la représentante de la Direction des affaires publiques de Netflix, les algorithmes n'interviendront pas non plus à l'avenir directement dans la création. En effet, Marie-Laure Daridan indique que l'entreprise dispose d'un chef de la production (*chief producer*). Elle affirme ainsi sa préoccupation de proposer désormais de la création originale et non plus des productions issues du travail computationnel de l'intelligence artificielle¹⁴. Par sa défense de la création, elle justifie la valeur symbolique de la culture. Tout se joue autour de la manière dont cette valeur symbolique est stratégiquement mise en avant pour le public, afin de produire un processus de consommation continue.

Ainsi se dessinent trois promesses de Netflix : celle de l'adéquation entre l'offre éditorialisée et la consommation, celle de la facilitation d'accès aux œuvres eu égard à un catalogue pléthorique et celle de la valeur culturelle des produits audiovisuels proposés. Cependant, ces promesses, qui justifieraient l'exploration du consommateur, ne correspondent pas tout à fait aux pratiques de Netflix qui associe à la recommandation basée sur les données d'usage une intervention humaine d'indexation visant à qualifier les contenus dans l'objectif de la création de métadonnées (Farchy et Denis, 2020, p. 164). Elles sont aussi et surtout circonscrites aux finalités économiques dans la mesure où même la modélisation des données a pour réel objectif de déterminer l'audience éventuelle d'un contenu sur le plan tant quantitatif que qualitatif. Cela vise à permettre une réduction des coûts financiers, par exemple en économisant sur la sortie d'un épisode pilote d'une série ou en adaptant en amont le budget dédié à une production suivant son public potentiel (Farchy et Denis, 2020, p. 135). Mais ces promesses d'exploration, relatives à l'accès aux productions audiovisuelles, engendreraient des attentes du consommateur par rapport à ce que Netflix

13. Propos de Guillaume Chaslot (ancien ingénieur chez YouTube et chercheur Mozilla Fellow en 2020) collectés lors de l'observation de l'événement « La réforme de l'audiovisuel en France : entre nouveaux acteurs et vieilles contraintes » (Université Paris 1, 3 mars 2020). Voir également concernant la comparaison entre Netflix et YouTube sur le site Web de la Fondation Mozilla. Mozilla, « Mozilla explains why does YouTube recommend conspiracy theory videos », <<https://foundation.mozilla.org/fr/blog/mozilla-explains-why-does-youtube-recommend-conspiracy-theory-videos/>> (consulté le 13 janvier 2022).

14. Chloé Delaporte rappelle à juste titre que les dirigeants de Netflix, pour justifier la proximité avec leur public, ont précédemment tenu un discours sur la manière dont une nouvelle version de la série *House of Cards* (BBC, 1990) aurait été produite. Les algorithmes de Netflix auraient permis d'entrecroiser l'intérêt pour les films réalisés par David Fincher avec ceux où figure l'acteur Kevin Spacey et le visionnage de la mini-série britannique *House of Cards* déjà présente dans le catalogue (Delaporte, 2018).

pourrait lui apporter en comparaison de la concurrence, telle que celles d'Amazon Prime Video ou Disney Plus¹⁵.

2. BRICOLAGE ET TRAVAIL DU PROFIL PAR LES ABONNÉS ET ABONNÉES

L'observation équipée offre un cadre adéquat pour étudier la manière dont s'articulent médiations techniques, éditoriales et symboliques. L'enjeu est de comprendre comment les abonnés et abonnées s'orientent dans le dispositif, quelle attention ils prêtent à la médiation algorithmique et quel rôle cette dernière joue dans leur pratique de consommation. Les observations nous ont offert un accès à l'expérience ordinaire des abonnés et abonnées grâce à notre invitation à décrire leur compte et leur pratique de recherche. Nous nous sommes ainsi concentrés sur la nature symbolique de la pratique de consommation en cherchant à nous placer au plus près des processus d'analyse et d'interprétation ordinaires des enquêtés et enquêtées.

Nous avons cherché à savoir ce qui se produit lors de la pratique de consommation du dispositif Netflix. Le compte de l'abonné ou l'abonnée est l'espace dans lequel cette consommation prend forme et est informée par les différentes médiations mises en œuvre par l'opérateur. Notre enquête pointe l'attention, voire le soin, que chaque abonné ou abonnée porte à son compte, ce qui se traduit concrètement par leur conscience de la production d'indices à l'attention de la plateforme en lien avec leur consommation.

2.1 À chacun son profil : le profil comme « machine à créer des traces », des goûts

L'opérateur propose, lors de la création d'un compte d'abonné, la constitution de différents profils pour les membres partageant le compte. Cette fonctionnalité se révèle très investie par les personnes enquêtées : la plupart nous ont d'emblée précisé que, lors de leur abonnement, elles ont immédiatement procédé à la création de leur profil et incité leur famille ou leurs amis et amies à faire de même. Cette démarche participe de la « fabrique de soi » au sein des industries médiatisantes (Gomez Mejia, 2016) et la prolonge dans un espace moins exposé, celui d'un compte d'abonné.

15. Hadopi et Conseil supérieur de l'audiovisuel, « La multiplication des services de vidéo à la demande par abonnement – Stratégies de développement et impact sur les usages », 9 mars 2021, p. 20, <https://www.hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2021_03_09_multiplication_services_video_demande_abonnement.pdf> (consulté le 13 janvier 2021).

Tous les enquêtés et enquêtées nous ont également souligné l'importance d'avoir des profils propres à chaque membre partageant un même compte. L'enjeu de la création de profils est double : d'une part, il y a une volonté de pouvoir inscrire sa pratique singulière dans un espace protégé, pour que celui-ci ne soit pas *brouillé* par les visionnages et les explorations des autres. D'autre part, il y a une volonté de pouvoir garantir une pratique de consommation individuelle continue en s'appuyant sur les fonctionnalités comme « la liste », « poursuivre avec le profil », téléchargement dans « mes vidéos ».

Certains enquêtés et certaines enquêtées soulignent le soin apporté à renseigner l'architexte (Souchier et Jeanneret, 1999) sur leurs goûts. Élodie [les prénoms ont été changés] nous indique : *« J'avais dit à ma mère "non là regarde bien, sinon tu vas te retrouver avec des mangas", elle déteste les mangas ! Et du coup, elle était très précise »*. Chloé invite les amis avec lesquels elle partage son compte à *« créer votre petit compte [c'est-à-dire un profil] à vous et parce que c'est individualisé. Et après vous faites ce que vous voulez »*. Remplir un formulaire correspond à une mise en écriture des goûts et de la pratique de consommation audiovisuelle et cinéphile. Cette activité s'inscrit dans une anticipation des futures médiations algorithmiques. Il est important de souligner que ce sont des traces de la pratique qui sont fournies en réponse au formulaire. Elles acquièrent un intérêt parce qu'elles sont reconnues comme saillantes. La force du dispositif est d'amener tout abonné et toute abonnée à exprimer ses goûts, à les rendre tangibles grâce à leur mise en écriture.

Se manifeste ici l'opérativité sociale et symbolique du dispositif de recueil mis en place par l'opérateur. Les processus de médiations mis en œuvre permettent un recueil régulier et continu d'indices des pratiques de consommation transformés en réel du goût des personnes. Ils s'inscrivent dans la perspective d'attribuer une valeur non négligeable aux propositions et à l'éditorialisation du catalogue dans l'espace du profil. Justine souligne ainsi que *« c'est vachement bien d'avoir des profils parce que ça va, pour chaque personne, s'adapter, peut-être recommander, en fonction du contenu regardé »*. L'abonné ou l'abonnée donne des signes qu'il sait qu'ils seront traités pour en produire de nouveaux. Ainsi, le profil, en tant que « machine à créer des traces » (Jeanneret, 2019), est analysé par les enquêtés et enquêtées comme un espace de médiations techniques auxquelles est prêtée une certaine efficience.

2.2 L'activité en lien avec la projection du travail algorithmique

Les médiations techniques, en particulier algorithmiques, ne sont pas mentionnées à proprement parler dans l'espace de l'écran. Il y a une conscience de leur présence sans que les enquêtés et enquêtées désignent des lieux précis où elles seraient identifiables concrètement. C'est une présence diffuse qui conduit Ihsane et Élodie à conclure l'entretien sur le fait que nous n'avons cessé de parler d'algorithmes de recommandation, quand bien même le mot n'a pas été prononcé. Notre enquête tend à souligner la réflexivité des consommateurs de services qui s'appuient sur des systèmes de recommandation automatisés : ils établissent des liens entre leur pratique de consommation et l'organisation éditoriale de leur profil. Très rapidement, les enquêtés et enquêtées nous racontent leur pratique actuelle, ce qu'ils font sur leur profil : ce qu'ils sont en train de regarder, les genres audiovisuels qu'ils aiment regarder, leur film préféré. Mais ils indiquent aussi que Netflix va les considérer comme des « indices » de leur pratique passée, pour reprendre les termes d'Yves Jeanneret (Jeanneret, 2019, p. 54).

D'un côté, notre guide d'entretien invite les enquêtés et enquêtées à lire et à décrire leur profil, en particulier les écrans qui le composent. Cela révèle qu'ouvrir son profil, c'est se livrer sans arrêt à une pratique d'interprétation des signes qui sont donnés à voir aux abonnés (du pourcentage de recommandation au nom de la catégorie, en passant par le positionnement de cette catégorie dans la liste). Nous pouvons reconnaître et mettre en évidence la teneur symbolique des pratiques dans le récit qu'en font les abonnés et abonnées (le plaisir à revoir le nom de certaines productions, ou encore la capacité à qualifier sémiotiquement certaines images, nous y reviendrons) et l'investissement réel que chaque abonné et abonnée développe sur son profil. De l'autre, les enquêtés et enquêtées y voient aussi la manière dont l'opérateur constitue des traces de leur pratique pour pouvoir s'en saisir et les traiter automatiquement : tout au long des entretiens, le lien est fréquemment établi entre ce qui a été vu et ce qui est suggéré. Ainsi les abonnés sont-ils pris dans le dispositif de traces que constitue leur profil.

Ce constat nous a permis de distinguer trois attitudes face à ce traitement indiciel de leurs pratiques significantes. Tout d'abord, plusieurs de nos enquêtés et enquêtées estiment que leur pratique régulière de visionnage suffit à constituer un ensemble de données de consommation qui produira des recommandations en correspondance. Dans cet esprit, ils n'ont pas

d'autre activité que celle de regarder des productions dans le catalogue et d'effectuer des recherches pour les choisir. Ensuite, quelques enquêtés et enquêtées choisissent de renseigner ponctuellement l'opérateur, notamment lorsque celui-ci les sollicite formellement (en particulier, lors de l'ouverture d'un cadre en vue d'une évaluation après le visionnage d'une production). Ils fournissent ces indications, qui constituent de petites tâches qui lui sont prescrites, sans régularité, mais qui le motivent à participer (Dujarier, 2008, p. 50-58). Enfin, parmi nos enquêtés, il y a ceux qui cherchent à « nourrir » l'algorithme de l'opérateur, comme nous l'indique Alban. Il s'agit de fournir des informations dans l'objectif de renforcer le pouvoir de recommandation de l'algorithme. Fanny précise notamment : *« J'ai mis le petit like qu'on peut mettre pour montrer que c'est ce que j'aimais comme contenu »* ; et Iona de nous répondre de façon similaire : *« Le pouce, je le fais quand quelque chose me plaît énormément ou quand quelque chose ne me plaît vraiment pas du tout, pour qu'on ne me propose pas des contenus qui ne me plaisent pas »*. Nous pouvons déceler ici un « travail » de la part de nos deux enquêtées. Comme le souligne Marie-Anne Dujarier, cette activité marque un investissement personnel pour participer à la production de l'environnement matériel et symbolique. Les quelques enquêtés et enquêtées qui cherchent à donner des indications les considèrent à partir de leur potentialité : elles peuvent jouer un rôle pour affiner les propositions futures. L'appréciation d'un contenu est dès lors constituée en une « logique d'indice » (Jeanneret, 2019, p. 204) par les abonnés et abonnées qui référencent leur pratique de jugement dans le dispositif.

Notre enquête souligne donc combien les consommateurs se positionnent dans leur pratique de consommation et de renseignement par rapport à cette médiation algorithmique : ils ne sont pas le jouet de ces médiations, ils les perçoivent et se positionnent par rapport à ce qu'ils en attendent en matière de personnalisation et d'efficacité dans la découverte de productions audiovisuelles.

2.3 Gestion du temps de recherche et mode d'exploration

En conséquence, si la principale force des algorithmes telle qu'elle est présentée par l'opérateur est censée être l'efficacité de l'appariement entre « offre large » et « goût ciblé » (ce dont nous reparlerons dans la dernière partie), les arguments liés à la rapidité des calculs jouent un rôle pour le consommateur de vidéos dans son rapport au temps, en particulier celui

qu'il consacre à la recherche de nouvelles œuvres à regarder. Mélanie, une autre enquêtée, nous disait en ce sens :

C'est ça qui est embêtant avec Netflix, et j'en ai discuté avec des potes à moi, c'est qu'on passe plus de temps à chercher un film qu'à visionner le film en lui-même [...]. Si je n'ai rien de précis, je cherche, mais je perds trop de temps [...]. Mais au final j'ai l'impression qu'il y a tellement de films, tellement de documentaires, tellement de séries sur ce type de plateforme que parfois on abandonne même le fait de regarder un film.

Ainsi, le rapport au temps de recherche était très présent dans le discours des enquêtés et enquêtées, qui se disent « agacés » de passer « trop de temps » à chercher de nouveaux contenus à visionner au sein de l'offre « pléthorique » que le dispositif est justement censé trier en fonction des goûts du consommateur.

On retrouve, à travers ce constat, un point largement traité par la littérature scientifique, particulièrement celle qui est inscrite dans une approche socioéconomique des plateformes culturelles. Elle montre que les « dispositifs de régulation de l'attention » (Mellet, 2017, p. 95) ont été élaborés précisément pour limiter la « frustration » liée à l'incapacité de « bénéficier totalement de ce catalogue sans limite » (Bourreau *et al.*, 2017, p. 108) – entendu que ce sentiment de frustration peut conduire à limiter l'utilisation du service lui-même. L'enjeu, au travers de la mise en place de tels dispositifs (algorithmiques, mais aussi esthétiques – design d'interface –, d'éditorialisation, etc.), est donc de « satisfaire, dit-elle, le consommateur pour qu'il reste client de la plateforme et ne décide pas de la quitter pour un concurrent » (*ibid.*, p. 109). Or, on voit bien se maintenir lors de nos entretiens un tel sentiment de frustration dans l'expérience de navigation. Une question reste donc en suspens : comment les abonnés s'y prennent-ils, eux-mêmes, pour limiter cette frustration, dans la mesure où les dispositifs précédemment évoqués échouent, au moins partiellement, à le faire ?

Nos observations ont alors permis de déceler un interstice entre ces deux pôles de tension (temps/recherche). Celui-ci participerait à l'émergence de « compétences » spécifiques, au moyen de « tactiques » propres qui se sont retrouvées de manière récurrente au cours des entretiens effectués. Justine nous expliquait par exemple : « Alors là, déjà que c'est difficile de choisir quoi regarder. [...] Rien qu'à l'image, il y a des choses qui ne m'intéressent pas du tout, ou je le sais directement. Quand je ne connais pas, c'est ça [l'image] qui va me guider. »

Est ainsi mis en œuvre un savoir d'analyse des images, par le développement de ce que nous nommons un « œil sémiotique », faisant que l'image va jouer un rôle ou non dans le processus de décision, donc de découvrabilité. Plusieurs enquêtées ont en ce sens indiqué y prêter une attention particulière, affirmant savoir « tout de suite » si la production va les intéresser ou non, par l'établissement d'un lien direct entre l'image et le genre du contenu, voire le scénario. Voici un autre extrait d'entretien illustrant ce point : *« C'est bête, mais quand je vois une image qui fait petit budget ou avec des effets mal faits et un graphisme pas beau et harmonieux, je me dis si déjà l'affiche, elle est pas belle, elle m'inspire pas, je me dis ça sert à rien, le film, je ne vais pas l'aimer »* (Élodie). Cette compétence, construite à partir de leur expérience et de leur culture visuelle et esthétique, de leur connaissance du catalogue et d'un affinement de leurs propres goûts, est ainsi mobilisée par les enquêtés et enquêtées dans un processus de sélectivité, qui peut d'ailleurs être conscientisé ou non. En effet, au cours des observations de leur pratique de navigation dans le dispositif, nous avons pu constater que même celles et ceux qui n'explicitaient pas directement dans leur exposé d'exploration le fait de s'appuyer sur les images pour sélectionner les contenus à visionner, ou au contraire à exclure d'emblée, mettaient constamment en pratique cette « compétence ».

Par ailleurs, cette dernière a conduit certains enquêtés et certaines enquêtées à prendre conscience de la « présence » des algorithmes, ou du moins à « voir » leur intervention dans les contenus suggérés :

Je me souviens quand j'ai regardé la série The Queen Gambit, après l'actrice, elle était en photo dans une série où elle avait un personnage mineur que j'avais déjà vu. Je me suis dit qu'ils l'avaient mise parce qu'en ce moment elle est très populaire avec cette série, et du coup sur tous les films ou séries où elle pouvait être, elle était partout en image de profil. Alors qu'une semaine avant ce n'était pas du tout elle qui était dessus. Et ça arrive. Quand je l'avais vue, ça m'avait beaucoup marqué, mais je sais que ça arrive d'autres fois aussi : regarder un contenu et la semaine d'après la photo d'une série plus bas elle a changé, en fonction de ce qu'on a regardé avant (Fanny).

Le lien est ici fait par l'enquêtée, qu'il soit réel ou non, entre contenus précédemment visionnés et propositions de visionnage futurs, où les secondes seraient donc adaptées par l'algorithme de recommandation, par l'image, en fonction des premières. Sans le savoir, l'enquêtée a ici repéré le travail algorithmique effectué par le système de personnalisation mis en place par la plateforme nommé Meson. Comme le mentionne

l'entreprise dans son Netflix Tech Blog, un des principaux objectifs de ce dernier est « d'augmenter la vélocité, la fiabilité et la répétabilité des expériences algorithmiques » pour, finalement, leur permettre d'afficher le slogan suivant : « Prédire ce que vous voulez regarder avant que vous ne le regardiez¹⁶ ». L'outil basé sur l'intelligence artificielle permet en effet de constamment actualiser les pages de navigation, et ainsi de personnaliser chaque interface en fonction de chaque utilisateur. C'est un élément que nous avons pu constater en visualisant les pages d'accueil de chacune des personnes avec qui nous nous sommes entretenus (titres et contenus de catégories, organisation de celles-ci, images/vignettes d'illustration des films et séries).

Ainsi, l'expertise déployée par les abonnés et abonnées dans l'interprétation des images, conduisant à une recatégorisation « mentale et immédiate » à l'intérieur même des catégories proposées par le dispositif, leur permet d'optimiser la recherche de nouveaux produits audiovisuels à visionner et ainsi de limiter un temps d'exploration perçu comme négatif. Cette forme de rationalisation dans le processus de recherche est par ailleurs renforcée par des modes d'exploration relativement « routinisés ». Karim, lorsque nous lui demandions de nous montrer la manière dont il s'y prenait pour chercher une nouvelle œuvre à regarder, nous expliquait par exemple :

Si on est en vacances, j'irai dans « Séries », mais si je n'ai pas beaucoup de temps, j'irai directement dans les « Films ». Si j'ai peu de temps, je regarderais un film assez court et qui me plaît, genre les trucs assez connus, et je pourrais clairement aller dans les « Plus gros succès Netflix », en général, ça ne rate pas, il va forcément y avoir un truc qui va me plaire.

Si les routines établies varient d'un enquêté à l'autre, il ressort bien de l'extrait d'entretien exposé ici une logique d'optimisation dans le processus d'exploration et de recherche (qui est plus adapté au contexte et au moment du visionnage) et qui, pour la plupart d'entre eux, se couple à une forte « habileté technique » dans leur navigation au sein de l'espace

16. D'un point de vue plus technique, voilà la définition qu'ils en donnent : « *To do this, we run a large number of machine learning (ML) workflows every day. In order to support the creation of these workflows and make efficient use of resources, we created Meson. Meson is a general purpose workflow orchestration and scheduling framework that we built to manage ML pipelines that execute workloads across heterogeneous systems. It manages the lifecycle of several ML pipelines that build, train, and validate personalization algorithms that drive video recommendations* ». Netflix Technology Blog, 2016, <<https://netflixtechblog.com/meson-workflow-orchestration-for-netflix-recommendations-fc932625c1d9>> (consulté le 29 septembre 2021).

contraint du textiel (Mayeur et Desprès-Lonnet, 2020), c'est-à-dire de l'écran qui peut être lu et sur lequel on peut agir.

Dès lors, l'attente formulée par le consommateur d'une forme d'immédiateté dans la recherche de nouveaux contenus (« passer le moins de temps possible à trouver un nouveau film ou une nouvelle série “intéressante” à regarder »), mise en regard avec la promesse qui lui est faite par l'opérateur d'avoir accès facilement à des productions audiovisuelles adaptées à ses goûts grâce à la performance de ses algorithmes de recommandation, soulève en dernier lieu la question de la « diversité » au sein des processus de découvrabilité.

3. DE L'APPARIEMENT À LA DÉCOUVRABILITÉ DANS LA CONSOMMATION AUDIOVISUELLE

Il s'agit là d'un point non négligeable dans les enjeux de découvrabilité. En effet, il nous semble que la question du temps (consacré à la recherche) est à mettre en relation avec la critique récurrente formulée dans les entretiens autour de l'idée suivante : « on me propose toujours la même chose ». Pointe ici le reproche, très largement traité par la littérature scientifique, selon lequel les algorithmes participent à restreindre la sérendipité, par une limitation des possibilités de découverte du « radicalement nouveau », comme le dit Vincent Bullich (2019, p. 285). L'algorithme ne proposerait en effet « que des contenus culturels ou informationnels liés aux parcours de navigation de chacun, parcours reconstruits à partir des mesures d'affichage. Il établirait ainsi une offre à partir du « toujours déjà là », et générerait un risque d'uniformisation de ce qui est proposé à chaque individu » (*ibid.*). Dès lors, un paradoxe est ici posé par les enquêtés et enquêtées : si le temps de recherche doit être limité, et si les algorithmes de recommandation proposent « toujours la même chose », quelle place reste-t-il à la découverte de contenus « différents », dans un catalogue pourtant promu comme varié, et comme favorisant ainsi la « diversité culturelle » (Bouquillion, 2020) ? En effet, toute prospection de productions audiovisuelles « autres », c'est-à-dire sortant de la « bulle » construite par les logiques de personnalisation (dans laquelle 78 % de nos enquêtés et enquêtées disent totalement ou partiellement « se reconnaître »), ne nécessite-t-elle pas, *de facto*, un temps relativement long d'exploration et de recherche ?

Cela soulève, en premier lieu, la question de l'appariement. La promesse est en effet la suivante : la plateforme dispose d'une variété et d'une diversité importantes d'œuvres dans son catalogue, que les algorithmes seront

capables de faire correspondre aux goûts des consommateurs, que ceux-ci soient « calculés » à partir des traces laissées ou à partir des indications données par les abonnés et abonnées. Mais, dès lors que ces derniers estiment « tourner en rond », émergent les limites de la stratégie ainsi déployée par la plateforme. Diversité produite ne veut pas dire diversité consommée, y compris lorsqu'une médiation algorithmique, aussi puissante soit-elle, intervient entre les deux. C'est du moins ce que montre l'enquête que nous avons menée. La récurrente incapacité pour les enquêtés et enquêtées à trouver un nouveau contenu à regarder, malgré les tactiques et routines établies, peut les conduire en dernier recours, nous l'avons vu, à rompre ce que Nicolas Auray et Bruno Vétel nomment « l'agir exploratoire », qui consiste à « se rendre disponible à des perturbations » (Auray et Vétel, 2013) dans les pratiques de consommation (vidéoludiques dans leur cas, audiovisuelles dans le nôtre). « L'exploration curieuse » en « régime d'abondance » (Auray, 2011), en quelque sorte prise en charge, du moins médiée par les algorithmes, atteint ici ses limites, pour se tourner alors vers un visionnage dit de « stock », par lequel le consommateur « va chercher quelque chose qu'il connaît et a stocké » (Beuscart *et al.*, 2019, p. 35). Le recours à la barre de recherche, pour avoir accès *directement* à des contenus prescrits par les amis, la famille ou certains comptes suivis sur les outils-marques que sont les réseaux sociaux numériques, ou à « Ma liste », espace-catégorie alimenté par les usagers eux-mêmes au cours de leurs navigations multiples, nous était présenté comme ultime tentative avant déconnexion de la plateforme¹⁷.

CONCLUSION

Promettre une performance et une accélération constante du traitement des données, donc des recommandations émanant des calculs, suscite, comme nous l'avons dit précédemment, une certaine « attente » chez l'utilisateur. Celle-ci n'est pas sans renvoyer aux avancées technologiques liées aux mégadonnées (*big data*) présentes dans les imaginaires autour d'une dite « intelligence artificielle », dorénavant capable d'étudier des masses de données toujours plus importantes, à une vitesse toujours

17. Du moins pour celles et ceux qui indiquaient utiliser la recommandation algorithmique comme source principale de découverte de nouveaux contenus, et qui représentent 60 % des enquêtés ayant donné une indication en ce sens, les 40 % restants déclarant mobiliser en premier lieu les autres formes de prescriptions mentionnées ci-dessus, que Sophie Maisonneuve nomme les « réseaux ou dispositifs de confiance interpersonnels » (Maisonneuve, 2019, p. 55) : famille/amis (25 %), RSN par certains comptes Twitter et Instagram (15 %).

croissante (imaginaires qui s'inscrivent dans des récits technologiques comme celui des « 5V » popularisé par Douglas Laney et censé représenter la performance des dispositifs techniques : volume, vitesse, variété, valeur, véracité ; Bourany, 2018). Au-delà de la revendication de la complexité technique du dispositif, capable de se mettre à jour constamment en fonction de multiples paramètres afin de personnaliser au maximum l'expérience utilisateur, notre perspective était bien ici d'examiner le rapport des usagers à ces dispositifs.

Finalement, ce rapport s'inscrit au cœur des enjeux techniques de la plateforme et des discours d'escorte qui en découlent, d'une part, et il pointe la teneur symbolique des pratiques telles que les enquêtes et enquêtées en témoignent, d'autre part. L'enjeu était d'interroger les « promesses » faites par l'opérateur à ses abonnés et abonnées et plus généralement aux consommateurs, dans ses discours promotionnels, à l'aune des pratiques, et de ce fait d'éviter de contribuer à la construction de leur « dimension idéologique », en pointant la « puissance et l'efficacité » des algorithmes (*ibid.*).

Cette démarche rejoint en ce sens les problèmes soulevés par l'ouvrage autour de la découvrabilité et de la diversité culturelle. Si grâce à notre recherche pointe la stratégie déployée par la plateforme de « mise en visibilité » (Bouquillion, 2020) de ses contenus par les systèmes de recommandation algorithmiques, émergent également ses limites. La capacité, vantée par Netflix, de favoriser la découvrabilité et la diversité culturelle, par la promesse de soumettre aux utilisateurs des œuvres « inédites », et en partie produites par l'opérateur, se heurte aux pratiques d'exploration, de tri et de sélection que nous avons pu observer. Ces pratiques ne se conforment finalement pas au « paradigme indiciaire », ce « “couteau suisse” de la justification des traces » (Jeanneret, 2019, p. 206), tant promu par les tenants d'une éditorialisation de l'offre guidée par le calcul des traces.

RÉFÉRENCES

- Auray, Nicolas et Bruno Vétel, « L'exploration comme modalité d'ouverture attentionnelle. Design et régulation d'un jeu freemium », *Réseaux*, n° 182, 2013, p. 153-186.
- Auray, Nicolas, « La consommation en régime d'abondance. La confrontation aux offres culturelles dites illimitées », *Revue française de socio-économie*, 2011/2, n° 8, p. 85-102.
- Bensamoun, Alexandra et Joëlle Farchy, *Mission intelligence artificielle et culture*, Paris, CSPLA, 2020.

- Beuscart, Jean-Samuel, Samuel Coavoux et Sisley Maillard, « Les algorithmes de recommandation musicale et l'autonomie de l'auditeur. Analyse des écoutes d'un panel d'utilisateurs de streaming », *Réseaux, La Découverte*, n° 213, 2019, p. 17-47.
- Bouquillion, Philippe, Bernard Miège et Pierre Moeglin, *L'industrialisation des biens symboliques : les industries créatives en regard des industries culturelles*, Grenoble, Presses universitaires de Grenoble, 2013.
- Bouquillion, Philippe, « Les stratégies de visibilité, le rôle des plateformes », *Annales des Mines, Enjeux numériques*, n° 10, juin 2020, p. 21-26, <<http://annales.org/enjeux-numeriques/2020/en-2020-06/2020-06-5.pdf>>.
- Bourany, Thomas, « Les 5 V du big data », *Regards croisés sur l'économie*, n° 23, 2018, p. 27-31.
- Bourreau, Marc, Michel Gensollen et François Moreau, « La consommation en régime d'abondance. Réactions des utilisateurs et stratégies des plateformes de streaming musical », dans Dominique Pasquier (dir.), *Explorations numériques : hommages aux travaux de Nicolas Auray*, Paris, Presse des Mines, 2017, p. 101-118.
- Bullich, Vincent, *Industrialisation, marchandisation et médiatisation des expressions : le modèle des plateformes numériques*. Contributions à la théorie des industries culturelles, Dossier pour l'obtention de l'habilitation à diriger des recherches, en sciences de l'information et de la communication, Grenoble, 2019.
- Catoir-Brisson, Marie-Julie et Laura Jankeviciute, « Entretien et méthodes visuelles : une démarche de recherche créative en sciences de l'information et de la communication », *Sciences de la société*, n° 92, 2014, p. 111-127.
- Coutant, Alexandre, Silva Drumond, Gabrielle Mota et Florence Millerand, « La production de l'utilisateur par les algorithmes de Netflix », *Les Enjeux de l'information et de la communication*, vol. 2, n° 19, 2018, p. 29-44.
- De Certeau, Michel, *L'invention du quotidien. 1. Arts de faire*, Paris, Gallimard, 1980.
- Delaporte, Chloé, « Dispositifs innovants, consommation créative ? Netflix ou la recommandation des contenus audiovisuels à l'ère de la prescription algorithmique », *Actes du Congrès de la Société française des sciences de l'information et de la communication*, MSH Paris Nord, juin 2018, p. 29-38.
- Dessinges, Catherine et Lucien Perticoz, « Les consommations de séries télévisées des publics étudiants face à Netflix : une autonomie en question », *Les Enjeux de l'information et de la communication*, 2019, n° 20, p. 5-23.
- Dujarier, Marie-Anne, *Le travail du consommateur. De Mac Do à eBay : comment nous coproduisons ce que nous achetons*, Paris, La Découverte, 2008.
- Farchy, Joëlle et Juliette Denis, *Culture des données : intelligence artificielle et algorithmes dans les industries culturelles*, Paris, Presses des Mines, 2020.
- Gomez-Mejia, Gustavo, *Les fabriques de soi ? Identité et industrie sur le Web*, Paris, MkF Éd., coll. Les Essais numériques, 2016.
- Hallinan, Blake et Ted Strichas, « Recommended for you : The Netflix Prize and the production of algorithmic culture », *New Media and Society*, 2016, p. 117-137.
- Hesmondhalgh, David et Amanda D. Lotz, « Video screen interfaces as new sites of media circulation power », *International Journal of Communication*, n° 14, 2020, p. 386-409.
- Jeanneret, Yves, *Critique de la trivialité : les médiations de la communication, enjeu de pouvoir*, Paris, Éditions Non Standard, 2014.
- Jeanneret, Yves, *La fabrique de la trace*, Londres, ISTE éditions, 2019.

- Maisonneuve, Sophie, « L'économie de la découverte musicale à l'ère numérique. Une révolution des pratiques amateurs? », *Réseaux*, vol. 1, n° 213, 2019, p. 49-81.
- Mayeur, Ingrid et Marie Després-Lonnet, « Aux origines du textiel. Entretien avec Marie Després-Lonnet », *Corela*, HS-33, 2020.
- Mellet, Kevin, « Le regard libre. Économies politiques de l'attention en régime numérique », dans Dominique Pasquier (dir.), *Explorations numériques : hommages aux travaux de Nicolas Auray*, Paris, Presses des Mines, 2017, p. 85-100.
- Paquiénéguy, Françoise, « Le glissement de la prescription dans les plateformes de recommandation », *Études de communication*, n° 49, 2017, p. 13-32.
- Souchier, Emmanuël et Yves Jeanneret, « Pour une poétique de "l'écrit d'écran" », *Xoana*, n°s 6/7, 1999, p. 97-107.
- Souchier, Emmanuël, Yves Jeanneret et Joëlle Le Marec (dir.), *Lire, écrire, récrire*. Paris, Bibliothèque publique d'information, coll. Études et recherche, 2003.

Autres références

- Hadopi et Conseil supérieur de l'audiovisuel, « La multiplication des services de vidéo à la demande par abonnement – Stratégies de développement et impact sur les usages », 9 mars 2021, <https://www.hadopi.fr/sites/default/files/sites/default/files/ckeditor_files/2021_03_09_multiplication_services_video_demande_abonnement.pdf> (consulté le 13 janvier 2021).
- Mozilla, « Mozilla explains why does Youtube recommend conspiracy theory videos », 8 juin 2021, <<https://foundation.mozilla.org/fr/blog/mozilla-explains-why-does-youtube-recommend-conspiracy-theory-videos/>> (consulté le 13 janvier 2022).
- Netflix Technology Blog, « Meson : Workflow orchestration for Netflix recommendations », 31 mai 2016, <<https://netflixtechblog.com/meson-workflow-orchestration-for-netflix-recommendations-fc932625c1d9>> (consulté le 29 septembre 2021).
- Netflix, « About Netflix », <<https://about.netflix.com/fr>> (consulté le 13 janvier 2021).
- ZDNet, « Comment Netflix compose ses algorithmes de recommandation », 26 avril 2017, <<https://www.zdnet.fr/actualites/comment-netflix-compose-ses-algorithmes-de-recommandation-39851718.html>> (consulté le 9 août 2021).

FAVORISER LA VISIBILITÉ DU PATRIMOINE CINÉMATOGRAPHIQUE PAR LES ALGORITHMES

UNE EXPÉRIMENTATION DE RECOMMANDATION DE CONTENU À CONTENU APPLIQUÉE AU CATALOGUE DE LA PLATEFORME TËNK

**Ève Givois¹, Éric Kergosien², Laure Bolka-Tabary³,
Samuel Gantier⁴ et Bernard Jacquemin⁵**

Comment rendre visible le patrimoine cinématographique ? Face à l'offre pléthorique et exponentielle des plateformes généralistes, les acteurs traditionnels de la valorisation du patrimoine cinématographique – cinémathèques, médiathèques, mais aussi distributeurs, programmeurs et même cinéphiles – sont confrontés à cet enjeu. Dans ce contexte, les acteurs de la médiation du cinéma patrimonial ne pourraient-ils pas bénéficier, à leur tour, de ces moyens pour toucher un public plus large ? Les technologies numériques dites intelligentes pourraient-elles améliorer la visibilité

1. Chargée de recherche postdoctorale au Centre Pompidou.

2. Maître de conférences à l'Université de Lille.

3. Maîtresse de conférences en sciences de l'information et de la communication à l'Université de Lille.

4. Maître de conférences en sciences de l'information et de la communication à l'Université Polytechnique Hauts-de-France.

5. Maître de conférences en sciences de l'information et de la communication à l'Université de Lille.

du patrimoine cinématographique ? En particulier, les dispositifs de recommandation algorithmique peuvent-ils aider à multiplier et optimiser les rencontres entre le public et les films de fonds, collections ou catalogues de films patrimoniaux⁶ ?

Pour explorer l'hypothèse d'une rencontre fructueuse entre les intermédiaires de la culture cinématographique et la pratique de la recommandation algorithmique de contenu à contenu pour l'amélioration de la visibilité du patrimoine cinématographique, nous nous appuyerons sur un cas particulier : celui de la plateforme de vidéo à la demande (VAD) et vidéo à la demande par abonnement (VADA) coopérative Tènk⁷. La plateforme, qui propose une sélection variée de films documentaires patrimoniaux et plus récemment réalisés, a constitué le terrain d'une recherche expérimentale menée en 2020-2021, dans le cadre du projet AlgoDoc⁸.

Nous commencerons par mettre en évidence les enjeux de visibilité du cinéma patrimonial sur le Web, pris entre une multiplicité d'acteurs et de pratiques de médiation, en nous concentrant sur l'espace français. La deuxième partie présentera les enjeux du déploiement d'un algorithme de recommandation dans un dispositif préexistant de médiation du cinéma documentaire, la plateforme Tènk. Il s'agira d'étudier le passage d'une recommandation éditoriale à une recommandation algorithmique, en relevant les contraintes de cette expérimentation. Dans une troisième partie, le processus d'expérimentation d'un système de recommandation algorithmique de films documentaires de contenu à contenu sera développé. Les métadonnées rassemblées et fabriquées pour expérimenter ce système seront mobilisées. La dernière partie examinera les limites de la démarche de recommandation automatisée pour faciliter la rencontre entre le patrimoine cinématographique et le public.

-
6. Le Centre national de la cinématographie (CNC) français définit les films patrimoniaux par l'ancienneté de leur première date de sortie en salles, supérieure à dix ans. Nous entendons plus largement dans ce chapitre les films de patrimoine comme l'ensemble des films pris dans un processus de patrimonialisation impliquant l'affirmation de leur appartenance à l'histoire du cinéma, du fait de leurs qualités ou de la position de leurs auteurs dans le champ cinématographique.
 7. Tènk, <<https://on-tenk.com/>>. Créée en 2016, dédiée au cinéma documentaire d'auteur, elle est implantée en Europe francophone (France, Belgique, Suisse, Luxembourg), au Canada et en cours de déploiement dans l'Union européenne et au Royaume-Uni.
 8. AlgoDoc (Algorithme de recommandation de films documentaires) est un programme financé en 2020 par l'I-Site et la Satt Nord. Il réunit un consortium recherche-industrie composé des laboratoires GERiCO de l'Université de Lille et DeVisu de l'Université Polytechnique Hauts-de-France, ainsi que Tènk et Spideo, société spécialisée dans la recommandation algorithmique de contenu audiovisuel. Spideo, <<https://spideo.com/>>.

1. ENJEUX DE LA VISIBILITÉ DU CINÉMA PATRIMONIAL SUR LE WEB

1.1 Mutations des acteurs et pratiques de médiation du cinéma patrimonial

L'activité de médiation du cinéma patrimonial dans l'industrie cinématographique a connu plusieurs mutations concomitantes ces dernières années en France.

« L'ivresse patrimoniale » du cinéma, qui s'amorce à la fin du ^{xx}^e siècle avec l'intégration du cinéma au sein du patrimoine national et international (Lindeperg et Szczepanska, 2017, p. 13), connaît une nette accélération au début du ^{xxi}^e. L'incroyable visibilité actuelle du cinéma patrimonial (Laurent, 2020b), en France et dans le monde, a d'abord été portée par les acteurs traditionnels de la médiation des films sortis du circuit commercial – cinémathèques, ciné-clubs, centres d'archives, mais aussi festivals dédiés. La dynamique a été accentuée par l'émergence et l'affirmation d'un marché du cinéma de patrimoine auquel contribuent producteurs et distributeurs (qui trouvent là un moyen d'exploiter leur capital), exploitants et diffuseurs (salles de cinéma, télévision et éditeurs de vidéo notamment) ou nouveaux festivals lui étant spécifiquement consacrés (Laurent, 2020a).

En parallèle, la numérisation et, dans une plus large mesure, le développement des technologies de l'information et de la communication (TIC) ont renforcé ce phénomène sur la durée. En premier lieu, les TIC contribuent à la mutation des pratiques traditionnelles de médiation du cinéma de patrimoine. C'est le cas notamment des festivals dédiés au cinéma patrimonial (Taillibert, 2015), du secteur traditionnel de la médiation du cinéma que constitue l'éducation à l'image (Taillibert, 2021), comme de la médiation des œuvres audiovisuelles éditées dans les bibliothèques et en médiathèque. Par ailleurs, les TIC font également émerger de nouveaux acteurs redéfinissant le secteur de l'industrie cinématographique et les enjeux patrimoniaux du cinéma. La médiation numérique des œuvres cinématographiques a ainsi connu une expansion continue depuis l'orée du ^{xx}^e siècle (Bullich, 2012). Les plateformes VAD/VADA, comprises à la fois comme un ensemble d'acteurs sociaux nouvellement constitués et comme outil technologique, composent désormais un intermédiaire central de distribution et diffusion du patrimoine cinématographique. Elles sont nombreuses dans ce nouveau paysage. Y dominent un petit nombre d'acteurs majoritaires internationaux généralistes, dont l'hégémonie est sortie renforcée de la crise sanitaire amorcée en 2020 et des confinements successifs. S'y maintiennent également plusieurs acteurs

minoritaires plus spécialisés (Direction des études, des statistiques et de la prospective, 2021, p. 45). Dans l'ensemble de ces plateformes, le cinéma patrimonial constitue une part non négligeable des catalogues de films mis en ligne (Danard, 2020). Les plateformes cinéphiliques⁹ contribuent à cette dynamique, mais, plus encore, participent du processus de patrimonialisation des œuvres qu'elles mettent à l'affiche. En mettant en avant l'aspect diachronique de leur programmation, en choisissant de mettre en avant un ensemble rétrospectif d'œuvres (Taillibert, 2020), elles contribuent à les consacrer.

Ces plateformes cinéphiliques s'inscrivent dans la logique de l'économie de la longue traîne (Benghozi et Paris, 2014) en ce qu'elles proposent à leurs utilisateurs des œuvres *a priori* peu rentables, qui, en étant diffusées en ligne, trouvent l'occasion d'être vues par un public autrement dispersé dans le monde physique. En ce sens, la dématérialisation de la médiation du cinéma patrimonial a pu être vue comme une manière opportune de toucher un public plus large et dispersé, voire de réaliser l'utopie de déconcentration des moyens de diffusion – ainsi, tout un chacun peut émettre, produire son propre média. Le phénomène charrie toutefois également son lot de menaces : ne soulève-t-il pas le risque d'une déshumanisation de la médiation, voire d'une perte de lien entre spectateurs ? Le risque est d'autant plus grand que la plupart de ces plateformes, du moins les plus populaires, dans le sillage de Netflix, mettent en avant leur utilisation d'algorithmes fondés sur l'expérience et les traces d'usages pour assurer à chaque utilisateur l'accès à une offre personnalisée.

1.2 Usage des algorithmes à des fins de recommandation

Les organisations proposant des biens culturels sur des plateformes disposant des dispositifs de recommandation algorithmique ont tendance à accentuer la visibilité des contenus les plus populaires (Dessinges et Perticoz, 2019) et à maintenir leurs utilisateurs dans des « bulles de filtres » hermétiques (Pariser, 2011), ou à proposer des contenus moins convaincants que les explorations autonomes de catalogue ou les prescriptions éditoriales (Beuscart, Coavoux et Maillard, 2019). De fait, ce type de recommandation est souvent opposé à une prescription éditoriale

9. Christel Taillibert a établi, en 2020, une liste de 17 plateformes cinéphiliques : Arte Créative et Boutique Arte, Benshi Studio, Bref Cinéma, Bretagne & diversité, CiNéMa Club, UPOPI et Ciel (Ciclic), Le Kinéscope, KuB, My French Film Festival, Nowave, Tènk, La Toile, Universciné, ainsi que Mémoire (Ciclic) et La Cinetek, dédiées spécifiquement au cinéma patrimonial (Taillibert, 2020, p. 140-141).

exigeante et éclairée : la première émanerait tendanciellement de plateformes généralistes et commerciales, tandis que la seconde serait plutôt mise en œuvre par des plateformes alternatives (Benghozi et Paris, 2016 ; Thuillas et Wiart, 2019a ; Thuillas et Wiart, 2019b). En outre, ce type de recommandation algorithmique prédictive personnalisée peut susciter la méfiance, par son opacité (Cardon, 2015) et sa normativité (Doueihi, 2013 ; Vayre, 2018). Dans ce contexte, les acteurs de la valorisation du patrimoine cinématographique seraient plus enclins à mettre de côté les dispositifs de recommandation algorithmique, qui ne répondent pas à leur besoin de rendre visibles des films en fonction de leur valeur patrimoniale ou de leurs qualités, toutes deux mesurées à l'aune d'un regard expert, et non de leur popularité ou des usages de leurs publics.

Pourtant, si les algorithmes basés sur les traces d'usage sont les plus répandus à des fins de personnalisation de la recommandation, d'autres peuvent être mobilisés. Les algorithmes de contenu à contenu, notamment, utilisent de leur côté uniquement les métadonnées¹⁰ associées en amont aux objets de la recommandation. Ils s'appuient sur l'ensemble des données techniques (supports, format, etc.), juridiques (visa, titulaire des droits, etc.), descriptives (titre, auteur, mais aussi genre, résumé, etc.) (Delcroix, 2018), plus ou moins élaborées, qui leur sont soumis, pour calculer la similarité entre les contenus et proposer une recommandation. En cela, ils diffèrent des algorithmes nourris par des données d'usage (profil et traces du comportement des utilisateurs, que ces derniers soient conscients de leur production – avis, notation – ou non – nombre de clics, parcours de navigation, etc.) (Farchy et Denis, 2020, p. 45). En évitant de recourir à une personnalisation de la recommandation, l'usage d'algorithmes de recommandation de contenu à contenu tend à limiter l'entretien des utilisateurs dans des bulles de filtres forgées à l'image de l'antériorité de leur navigation. En outre, l'explicitation de la recommandation est facilitée par l'affichage simultané de la source du calcul algorithmique (le contenu de départ) et de l'objet de la recommandation (le contenu prescrit).

Or, les acteurs de la valorisation du patrimoine cinématographique, en tant qu'intermédiaires de la culture cinématographique (Jeanpierre et Roueff, 2014), disposent bien souvent des connaissances, compétences et infrastructures indispensables pour élaborer des métadonnées forgeant

10. Les métadonnées peuvent être définies comme les « informations décrivant un contenu [...] précis selon des caractéristiques objectives ou plus ou moins construites », selon Farchy et Denis, 2020, p. 42.

pertinemment la recommandation algorithmique. Ainsi les professionnels des médiathèques, archives ou cinémathèques peuvent-ils mettre leur culture professionnelle de la documentation et de la prescription des biens culturels au service de l'enrichissement des métadonnées, quand celles-ci ne sont pas déjà disponibles au sein de leurs établissements. De leur côté, les distributeurs ou programmeurs qui travaillent au sein de plateformes VAD/VADA cinéphiliques, forts de leur expérience d'éditorialisation, ont l'expertise nécessaire pour qualifier pertinemment les films patrimoniaux qu'ils mettent en ligne, et participer à leur indexation. La convergence de ces différentes pratiques et savoir-faire ne permettrait-elle pas d'élaborer des outils algorithmiques plus humains, utiles et accessibles à chacun de ces acteurs ?

2. UN ALGORITHME DE RECOMMANDATION HUMAIN ?

Pour imaginer un tel dispositif de recommandation, revenons d'abord sur la fonction qui lui serait attribuée, profondément humaine : il s'agit de reproduire le mécanisme de recommandation tel qu'il est pratiqué par les professionnels de la médiation du cinéma, c'est-à-dire de mise en liens des films entre eux. La création de ces liens est l'enjeu de toute pratique éditoriale mise en œuvre par les différents acteurs de la médiation cinématographique. Ils peuvent se nouer dès le moment de la création des films ou de leur acquisition par les producteurs, distributeurs et diffuseurs. À l'autre bout de la chaîne, ils sont aussi tissés par les professionnels de la médiation audiovisuelle que sont les conservateurs, médiathécaires et documentalistes audiovisuels lors de la constitution de leurs collections ou de l'acquisition de leurs fonds. Ainsi, Jean-Yves de Lépinay, président de l'association Images en bibliothèque¹¹, considère que la mission des médiathèques est « de créer non seulement avec les usagers, mais aussi des liens entre les œuvres¹² ». Les plateformes VAD/VADA elles aussi, par leur programmation, contribuent à forger des liens entre les films.

11. Créée en 1989, elle a pour mission « d'affiner la culture cinématographique des bibliothécaires, mais aussi de faciliter le repérage et l'accès aux films de tous les genres » (Du Val, 2017). Elle structure depuis un réseau de structures adhérentes pour promouvoir la médiation du cinéma en bibliothèque, ainsi que la formation des professionnels et l'éducation à l'image. Voir Images en bibliothèques, « À propos », <<https://imagesenbibliotheques.fr/a-propos>>.

12. Jean-Yves de Lépinay (président d'Images en bibliothèque), « Quelle place et quelles missions pour les bibliothèques ? Appréhender les différentes offres et l'évolution des pratiques en médiathèque », *VaD et ressources numériques*, formation professionnelle en ligne organisée par l'association Images en bibliothèque à destination de bibliothécaires et médiathécaires, 29 avril 2021.

2.1 De l'éditorialisation à la recommandation algorithmique : le cas de Tènk

Parmi l'ensemble des plateformes VAD/VADA actuellement en ligne, nous nous intéresserons tout particulièrement ici à Tènk, plateforme proposant à ses abonnés un catalogue de films documentaires d'auteurs, dont une part de films patrimoniaux : 36 % des 2200 titres programmés depuis les débuts de la plateforme en 2016 étaient sortis depuis plus de dix ans au moment de leur première mise en ligne. Le projet à l'origine de Tènk, avant que ses créateurs ne décident de la création d'une plateforme, était de constituer une télévision dont le flux serait entièrement dédié au cinéma documentaire (Lesaunier, 2019). Il s'agissait ainsi pour les documentaristes, programmeurs du festival des États généraux du film documentaire et autres passionnés du genre se rassemblant en coopérative, de donner au cinéma documentaire d'auteur son autonomie envers les chaînes de télévision pour programmer, c'est-à-dire sélectionner et faire des liens entre films documentaires. Aujourd'hui, la plateforme propose à ses abonnés une sélection de sept films par semaine en ligne pendant deux mois, ainsi qu'un large catalogue d'œuvres en VAD. Son activité de médiation passe par une programmation incarnée, préparée par un réseau de bénévoles sur la base de leur culture individuelle et subjective du cinéma documentaire, validée par l'équipe de direction artistique. Le cœur d'activité de la plateforme est ainsi l'activité d'éditorialisation, soit un processus de choix et valorisation de contenus provoqué par leur numérisation (Bachimont, 2007 ; Epron et Vitali-Rosati, 2018). Pour cela, l'équipe de Tènk mobilise un savoir-faire s'inscrivant dans la continuité des activités de programmation événementielle et de distribution traditionnelle du cinéma. Après cinq années d'existence, le nombre d'œuvres ayant déjà été diffusées par la plateforme a considérablement augmenté : « Le catalogue est passé de 7 films par semaine à environ 700 films en [VAD]. Aujourd'hui, permettre aux utilisateurs de Tènk de s'y retrouver face à une offre de 700 films est primordial¹³ ».

Sans que le catalogue soit démesuré, ce changement d'échelle encourage l'équipe à mettre en place trois types d'actions de recommandation : poursuivre la recommandation éditoriale, en continuant à solliciter des programmeurs bénévoles pour proposer des associations de films

13. Entretien avec Pierre Mathéus et Samuel Gantier, mené par Némésis Srour (2021), « Can Algorithms be Ethical? Behind Machines, a Human-Centered Approach », *Rumo* : <<https://rumo.co/can-algorithms-be-ethical-behind-machines-a-human-centered-approach/>>, traduction des auteurs.

originales ; introduire de la recommandation sociale, en permettant aux abonnés de partager leurs coups de cœur auprès de leurs contacts ; et enfin développer une recommandation algorithmique (Gantier, 2020). Le déploiement d'un algorithme est ainsi pensé comme une façon de pallier les limites de la mémoire humaine des programmeurs de la plateforme et de concevoir une mémoire collective de la programmation. Il nécessite toutefois de mettre en œuvre un processus d'objectivation et de formalisation des critères de sélection éditoriale par l'indexation systématique du catalogue. Ceux-ci sont multiples, mais ne se limitent pas aux auteurs, aux périodes ou lieux de production, ou encore aux thématiques abordées. Ils s'appuient bien souvent sur les procédés sociotechniques mobilisés par les auteurs de ces films, leurs partis pris de réalisation, autrement dit leurs dispositifs cinématographiques.

2.2 Enjeux éthiques de la recommandation algorithmique

Cette expérimentation d'un dispositif de recommandation algorithmique n'est pas sans poser de questions éthiques, liées à la fois aux contraintes qui pèsent sur les différents partenaires concernés, et à la question plus large de l'usage d'algorithmes pour proposer des recommandations de biens culturels.

Tènk envisage de déployer un tel dispositif à la condition qu'il réponde à sa ligne éditoriale. La plateforme affiche ainsi la volonté de continuer à se positionner en opposition avec les acteurs majoritaires de la VAD, en adoptant une position conforme à ses valeurs coopératives. Elle tient à inscrire la recommandation algorithmique dans la continuité de ses pratiques de programmations collectives, subjectives et incarnées. En cohérence avec cette position, l'algorithme de contenu à contenu semble être adapté, puisqu'il permet à l'équipe de programmeurs de garder la main sur la fabrication des données nourrissant la recommandation. La technologie permet de prendre en considération des aspects subjectifs, humains, intelligents, de la recommandation, dès lors que ceux-ci sont décrits dans les métadonnées.

De son côté, la société Spideo, qui met à disposition du projet une API dédiée à la recommandation, Rumo, affiche une volonté d'améliorer significativement la transparence de son dispositif. L'enjeu pour l'entreprise, dont l'activité était jusqu'à peu concentrée sur l'analyse sémantique de contenus et la prestation de services documentaires (Farchy et Denis, 2020, p. 168 ; Gantier, Givois et Kergosien, 2020), est d'investir le marché de la

recommandation sans proposition de création de nouvelles métadonnées. Sa participation au projet lui permet à la fois de tester sa nouvelle API et d'accélérer son développement. Elle participe également plus largement d'une stratégie d'amélioration de la compréhension des mécanismes de recommandation algorithmique, dans l'idée de s'attirer la confiance de ses clients actuels et futurs dans ses outils.

Ce souci de transparence, aux implications commerciales dans le cas de l'entreprise Spideo, constitue toutefois, de manière plus générale, un enjeu majeur de l'éthique des algorithmes, car l'« opacité peut être vue comme un facteur favorisant les manquements ou les risques éthiques » (Simonnot, 2018 ; Adary et Domenget, 2017). À ce titre, l'utilisation d'un algorithme de recommandation de contenu à contenu semble encore s'imposer pour accroître la transparence dans la mesure où il facilite l'explicitation de l'opération computationnelle, en plus d'éviter l'intrusion dans les données personnelles des utilisateurs.

Préférer recourir à des algorithmes de recommandation centrés sur le contenu, plutôt qu'à d'autres mobilisant les traces d'usage des utilisateurs, présente l'avantage de ne pas s'affranchir d'une réflexion sur les modalités de fabrication de la donnée. Celle-ci n'est pas appréhendée comme la trace d'un « réel » qui lui préexiste (Bonenfant, Crémier et Lafrance St-Martin, 2018), mais comme un produit construit socialement, en fonction de besoins préalablement identifiés dans l'optique de représenter une dimension spécifique des œuvres afin de produire un effet bien particulier. Les métadonnées associées aux films à recommander sont construites sciemment, à dessein de recommandation, et non pas récoltées, sorties de leur contexte de production, malgré les usagers qui les produisent – comme le sont les traces d'usage. Contrairement à ces dernières, dont la production n'est pas socialement considérée comme un travail (Tiffon, 2017), les métadonnées dédiées à la recommandation de contenu à contenu sont élaborées par des travailleurs rémunérés à cette fin. Les œuvres sont documentées par leurs médiateurs traditionnels (et non par des contributeurs bénévoles d'une indexation folksonomique¹⁴) qui mettent ainsi à profit leurs compétences professionnelles – loin de l'activité de micro-travailleurs du clic (Casilli, 2019 ; Le Ludec, Wahal, Casilli et Tubaro, 2020).

14. Le terme a été forgé par la combinaison de *folk* (usagers) et *taxonomy* (règles de classification) par Thomas Vander Wal (2004) pour « désigner le phénomène d'indexation des documents numériques par l'utilisateur » (Le Deuff, 2006).

3. EXPÉRIMENTATION D'UNE RECOMMANDATION AUTOMATISÉE

Pour éprouver la mise en œuvre d'un tel algorithme de recommandation de contenu à contenu et, en particulier, son humanité, nous avons eu recours à l'expérimentation. Il s'agissait notamment de déterminer à quelles conditions l'utilisation d'un algorithme de recommandation peut pertinemment faciliter la découvrabilité par le public du patrimoine cinématographique documentaire.

3.1 « Mettre en données » les formes des films documentaires pour mieux les recommander

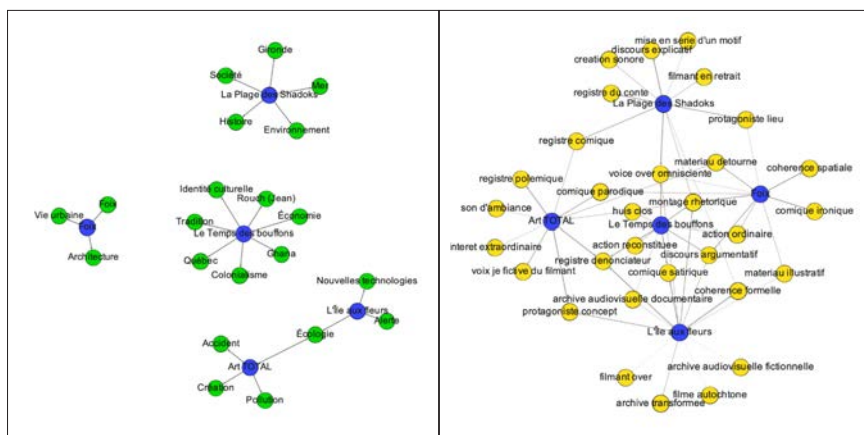
La première étape de notre expérimentation a consisté à objectiver les critères de rapprochement éditorial entre films mis en œuvre par l'équipe de Tënk, afin de transposer la démarche de recommandation humaine dans un outil algorithmique. En d'autres termes, pour se voir recommandées au moyen d'un algorithme, les œuvres de cinéma documentaire doivent, à leur tour, être passées au crible d'une « mise en données » (Menger, 2017 ; Vayre, 2017 ; Bullich et Clavier, 2018 ; Flichy, 2021).

Chez Tënk, les rapprochements entre films sont établis par les programmeurs moins en fonction des thématiques ou du genre (catégorie sans grande significativité lorsqu'on l'applique au seul cinéma documentaire, voir Kilborne, 2008) que sur des critères plus formels, davantage liés à la démarche d'auteurs et aux dispositifs choisis pour réaliser leurs films. La création d'un outil dédié à l'indexation de cet aspect des films documentaires a permis d'envisager la mise en place d'une recommandation automatisée se basant sur les aspects formels et sociotechniques, et non plus thématiques¹⁵. La collecte d'une connaissance experte de la variété de dispositifs mobilisés pour la réalisation de films documentaires, menée auprès des équipes éditoriales de Tënk, de chercheurs en études cinématographiques et au moyen de la littérature spécialisée, a donné naissance à un thésaurus des dispositifs de réalisation du cinéma documentaire (Givois *et al.*, 2021). Ce vocabulaire contrôlé et structuré, dont les termes peuvent être liés entre eux par des relations sémantiques d'équivalence (synonymie ou association) ou hiérarchiques (spécificité, généralité), a été mobilisé pour caractériser un ensemble de 177 films programmés par Tënk depuis sa création.

15. Voir la V1 du « Thésaurus des procédés filmiques du cinéma documentaires », publié sur *OpenTheso*, <<https://opentheso.huma-num.fr/opentheso/?idt=th228>>.

Au terme de cette démarche, il apparaît que les rapprochements entre films selon les dispositifs de réalisation du cinéma documentaire, indexés systématiquement sur un corpus restreint, s'avèrent plus riches que les rapprochements faits selon des métadonnées thématiques préexistantes (Gantier, Givois et Kergosien, 2020). La figure 10.1 permet de comparer les liens entre cinq films à la tonalité satirique et comique qui pourraient justifier de les voir recommandés entre eux : *Foix* (Luc Moullet, 1994), *La Plage des Shadoks* (Pierre Andrieux et Fabrice Marache, 2016), *Le Temps des bouffons* (Pierre Falardeau, 1985), *Art TOTAL* (Pierre Excoffier et Gwenn Pacotte, 2000) et *L'Île aux fleurs* (Jorge Furtado, 1989). Leur indexation thématique crée un seul lien entre deux d'entre eux : *Art TOTAL* et *L'Île aux fleurs*, qui traitent de sujets écologiques. Les trois autres sont complètement disjointes. À l'inverse, l'indexation des dispositifs de réalisation multiplie les liens entre ces cinq films : ainsi, quatre d'entre eux sont montés autour d'un discours (montage rhétorique), trois ont une voix hors champ omnisciente, trois autres relèvent d'un discours argumentatif, trois autres encore d'un registre dénonciateur. Ils ont en commun également un registre comique, que celui-ci soit parodique, ironique ou satirique.

FIGURE 10.1
Liens entre films différents selon le mode d'indexation



Légende : Le graphique propose une comparaison des liens entre films produits par l'indexation thématique (à gauche) et l'indexation des dispositifs de réalisation (à droite). Il est généré avec Gephi, ForceAtlas2 en mode LinLog avec déchevauchement et dissuadement des concentrateurs (*hubs*) puis ajustement des labels.

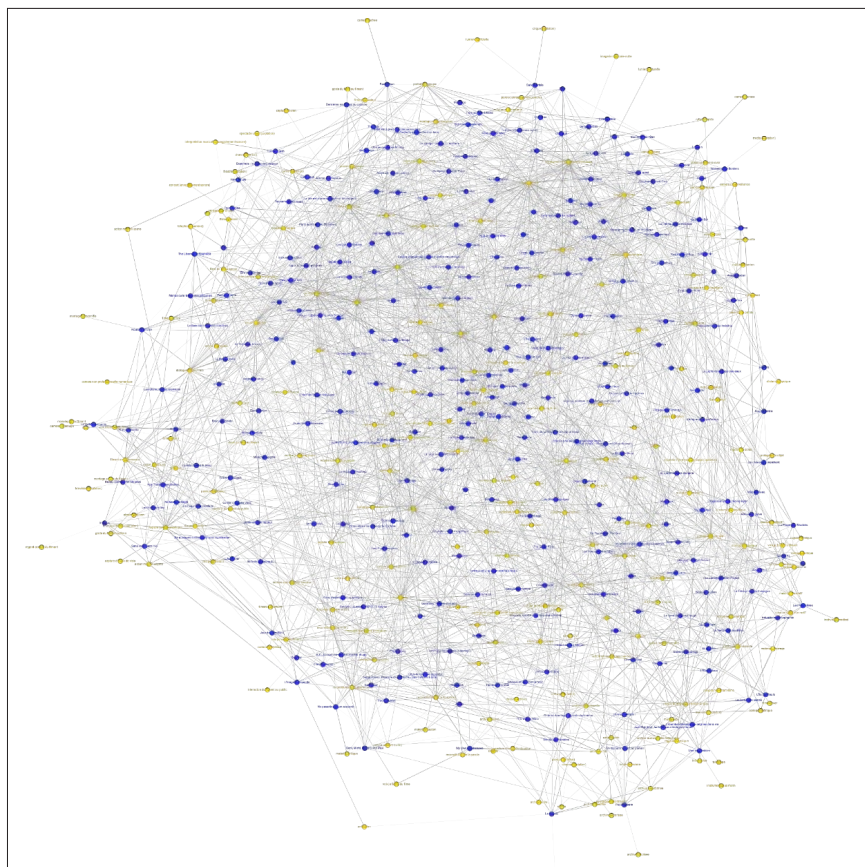
Ces premiers résultats permettent de confirmer l'hypothèse de la pertinence d'une indexation alternative à l'indexation thématique, dès lors que celle-ci est menée à partir d'un vocabulaire construit en fonction des particularités du catalogue de Tënk. En établissant des liens inexistantes jusqu'à lors entre films, nous pouvons déduire que le jeu de métadonnées ainsi établi rend possible l'amélioration de la visibilité des films proposés dans le catalogue de Tënk.

Toutefois, la démultiplication des liens entre films ne constitue pas une condition suffisante pour proposer aux utilisateurs de la plateforme des recommandations satisfaisantes. Encore faut-il que, de l'ensemble de ces nouveaux liens, ressortent bien les plus pertinents aux yeux des programmeurs dont le dispositif algorithmique cherche à imiter la recommandation. La figure 10.2, dans laquelle figurent les 177 films indexés et les termes décrivant leurs dispositifs de réalisation associés, rend prégnant l'enjeu de la hiérarchisation des liens entre films selon leur pertinence. La multitude de liens créés par l'indexation des dispositifs y apparaît fortement entremêlée.

Pourtant, si l'on zoome sur ce graphique à la recherche des cinq films au ton comique et satirique présentés dans la figure 10.1, on les retrouve bien à proximité (figure 10.3), au milieu d'autres avec lesquels ils partagent une forme de discours argumentative et un montage rhétorique (*Aujourd'hui d'Esther Choub*), une caractéristique formelle telle que l'intervention du réalisateur ou de la réalisatrice en voix hors champ (*La Femme aux cent visages* de Jean-Daniel Pollet) ou bien le détournement de matériaux de leur sens initial (*The Centrifuge brain project* de Till Novak).

La principale fonction de l'outil algorithmique mis à disposition par Spideo, Rumo, est de faire le tri des liens entre films formés par l'indexation. Il propose, pour chaque film, une liste de 30 autres ordonnés selon un score calculé au moyen d'une mesure de similarité cosinus.

FIGURE 10.2
Liens entre films formés par l'indexation
des dispositifs de réalisation



Légende: Les nœuds correspondant à des films sont colorés en bleu, les descripteurs en jaune.
Le graphique est généré avec Gephi, ForceAtlas2 en mode LinLog avec déchevauchement
et dissuadement des concentrateurs (hubs), avec l'indexation de 177 films du catalogue de Tènk.

3.2 Évaluer la pertinence de la recommandation algorithmique

Le fruit de cette recommandation soulève toutefois beaucoup d'interrogations. Ces liens sont-ils susceptibles d'amener les utilisateurs à découvrir des contenus à la fois nouveaux pour eux et similaires au film de départ ? Plus spécifiquement, sont-ils à la hauteur d'une recommandation éditoriale proposée par un programmeur ? Ces questions soulèvent celle de l'évaluation de la recommandation, de ses buts et de ses moyens.

Les buts, d'abord. L'évaluation de notre expérimentation a été menée dans deux directions différentes. Premièrement, nous avons cherché à distinguer les convergences et divergences entre recommandation éditoriale et algorithmique en les comparant. Deuxièmement, pour nous assurer que les métadonnées soumises à l'algorithme soient optimisées à des fins de recommandation, plusieurs formats ont été testés.

Les moyens, ensuite. Pour suivre ces pistes d'évaluation de la recommandation, plusieurs outils d'appréhension, visuelle et quantitative, ont été mobilisés. Le tableau de bord de Rumo n'apportant pas de vision globale de la recommandation et imposant un examen film à film, les outils de représentation de réseaux lui ont été préférés.

3.2.1 Buts de l'évaluation

Un des enjeux de l'implémentation d'un algorithme pour recommander des films entre eux sur la plateforme Tënk étant d'automatiser une tâche éditoriale humaine, son évaluation ne peut faire l'impasse d'une comparaison entre recommandation algorithmique et recommandation éditoriale. Cette comparaison implique plusieurs défis. Elle nécessite d'abord la récolte d'une recommandation éditoriale directe auprès d'un panel à la culture du cinéma documentaire assez riche et homogène pour que ses membres aient vu les mêmes films et qu'il soit possible de croiser leurs suggestions de recommandations. Dans cette perspective, huit spécialistes du cinéma documentaire (programmeurs de Tënk, chercheurs et formateurs en études cinématographiques, tous également auteurs de films documentaires) ont été consultés. Un premier corpus de 112 films leur a été soumis, en leur demandant les liens qu'ils peuvent faire entre chacun de ces films selon leurs caractéristiques formelles et discursives (voir Gantier, Givois et Kergosien, 2020). Ce recueil a permis de relever quelques liens forts entre films, des liens confirmés par au moins trois experts. Parmi eux, cinq familles de films fréquemment associés ont pu être recensées. Au terme de l'exercice, les limites de la démarche ressortent clairement : quatre

films sur 112 avaient été vus par les huit experts sollicités. Ce dispositif d'évaluation de la recommandation se trouve ainsi confronté à une limite d'échelle, celle-là même qui pousse à recourir à un algorithme pour générer des recommandations.

Parallèlement, la seconde piste d'évaluation a consisté à faire varier légèrement les métadonnées soumises à l'algorithme (nombre de descripteurs par film, avec ou sans pondération, avec ou sans prise en considération des liens sémantiques entre termes descripteurs du thésaurus) pour observer leurs effets sur la recommandation et déterminer quel format de métadonnées amenait les résultats les plus intéressants. Après la computation algorithmique, nous nous sommes également demandé s'il était possible d'établir un seuil de pertinence du score de la mesure cosinus calculée par Rumo¹⁶, ou si l'on pouvait considérer qu'au-delà d'un certain ordre, dans la liste des recommandations proposées, la pertinence des films proposés n'était pas moindre.

3.2.2 Moyens de l'évaluation

Pour mettre en œuvre ces comparaisons, les outils de visualisation et d'analyse de réseaux présentent l'avantage de donner une vision globale des résultats du calcul algorithmique.

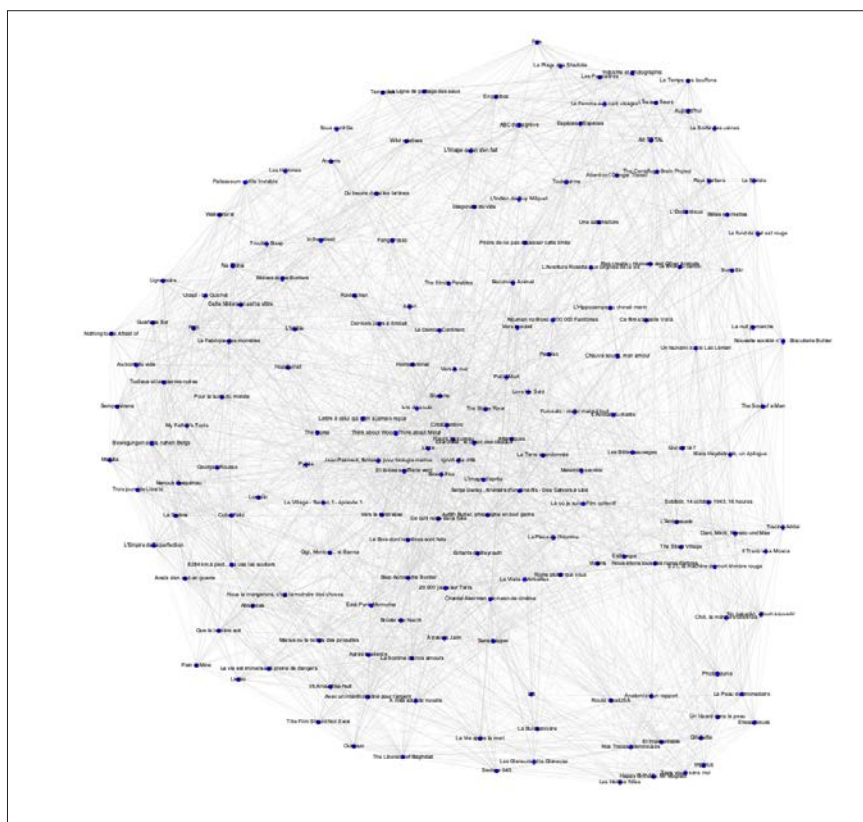
La visualisation du réseau formé par la recommandation sous forme d'un graphique nœuds-liens (figure 10.5), au même titre que la représentation des liens entre films et descripteurs (figure 10.2), n'est pas des plus lisibles. Tout au plus pouvons-nous repérer le positionnement des films dans le plan, repérer ceux qui semblent faire le lien entre plusieurs ensembles de films recommandés entre eux (au centre du graphique) et d'autres qui semblent former des sous-ensembles plus isolés (à la périphérie). Mais les liens, bien qu'ils soient proportionnels au score de similarité cosinus, se confondent et se superposent.

Pour dépasser ces limites et mettre en œuvre la comparaison entre recommandations humaine *et* algorithmique, nous nous sommes donc tournés vers une autre forme classique de représentation des réseaux, la matrice d'adjacence. Celle-ci place chaque sommet à la fois en ligne et en colonne d'un tableau à double entrée. Si deux sommets sont reliés, la case à

16. Sur les modalités du recueil de la recommandation éditoriale, mais aussi sur les enjeux de l'utilisation de cette mesure cosinus, dite cosinus de Salton, voir Gantier, Givois et Kergosien, 2020.

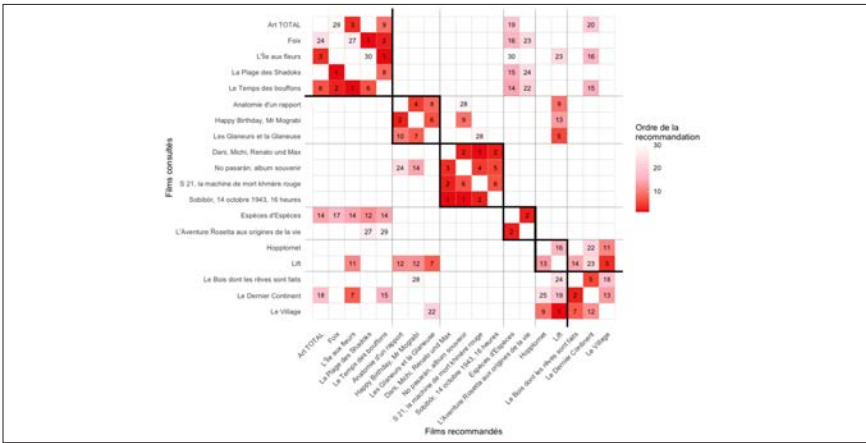
l'intersection de leur ligne et de leur colonne est remplie d'une valeur numérique continue caractérisant le lien. Malgré les difficultés d'appréhension de l'outil, moins intuitif que la représentation en graphique nœud-lien, la matrice de contingence présente notamment l'avantage de rendre plus visibles les absences de connexion et, du même coup, de faciliter l'identification de communautés de sommets, dès lors qu'un réordonnancement des lignes et colonnes en fonction des liens est opéré (Henry et Fekete, 2008). Ainsi, pour comparer recommandation éditoriale *et* algorithmique, les films qui avaient été recommandés entre eux par trois membres du panel ont été ordonnés en cinq groupes, reconstituant les familles de films préalablement identifiées.

FIGURE 10.5
Représentation en réseau des liens algorithmiques
entre 177 films du catalogue de Tènk



Légende: La représentation s'appuie sur une indexation des dispositifs cinématographiques pondérés selon une échelle par paliers (1, 3, 5, 8, 10).

FIGURE 10.6
Représentation en matrice des liens entre films proposés par le panel et confirmés par l’algorithme (pondération log2)

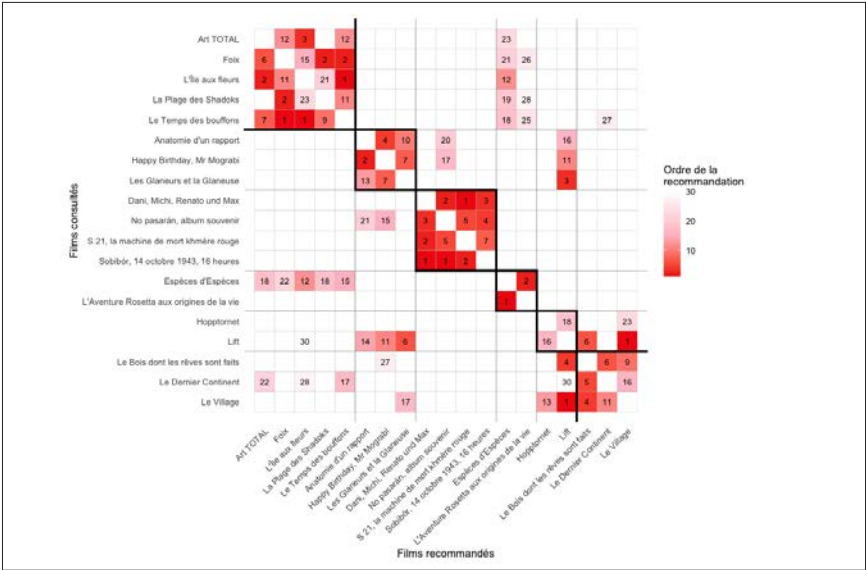


Légende: Cette matrice d’adjacence présente les liaisons algorithmiques entre films du corpus associés entre eux par au moins trois des membres du panel de spécialistes, s’appuyant sur une indexation des dispositifs cinématographiques pondérés selon une échelle logarithmique de base 2. Les liens sont caractérisés par l’ordre du film dans la recommandation proposée par Rumo.

Cette matrice d’adjacence laisse apparaître une forme d’équivalence entre recommandation algorithmique et recommandation éditoriale : les films reliés entre eux par les experts se trouvent bien recommandés ensemble après un calcul de similarité cosinus, les blocs centraux de la matrice étant bien remplis de liens, plus ou moins forts selon les groupes. Des liens extérieurs aux blocs centraux sont toutefois à noter.

La visualisation en matrice d’adjacence a permis également de comparer les formes prises par la recommandation après avoir fait varier le nombre de descripteurs pris en considération ou leur pondération notamment. Les figures 10.6 et 10.7 montrent ainsi l’évolution de la recommandation en fonction de la pondération : la figure 10.6 a été générée à partir de métadonnées pondérées suivant une échelle logarithmique de base 2, tandis que la figure 10.7 avait pour base des données pondérées suivant une échelle par paliers (détaillés dans le tableau 10.1).

FIGURE 10.7
Représentation en matrice des liens entre films proposés par le panel et confirmés par l’algorithme (pondération par paliers)



Légende : Cette matrice d’adjacence présente les liaisons algorithmiques entre films du corpus associés entre eux par au moins trois des membres du panel de spécialistes, s’appuyant sur une indexation des dispositifs cinématographiques pondérés selon une échelle par paliers (1, 3, 5, 8, 10). Les liens sont caractérisés par l’ordre du film dans la recommandation proposée par Rumo.

TABLEAU 10.1
Deux échelles de pondération testées et évaluées dans les figures 10.5 et 10.6

CATÉGORIE DES DESCRIPTEURS	PONDÉRATION SELON UNE ÉCHELLE LOG2	PONDÉRATION PAR PALIERS
Descripteurs n°s 1 à 3	16	10
Descripteurs n°s 4 et 5	8	8
Descripteurs n°s 6 à 8	4	5
Descripteurs n°s 9 et 10	2	3
Descripteurs n°s 11 et suivants	1	1

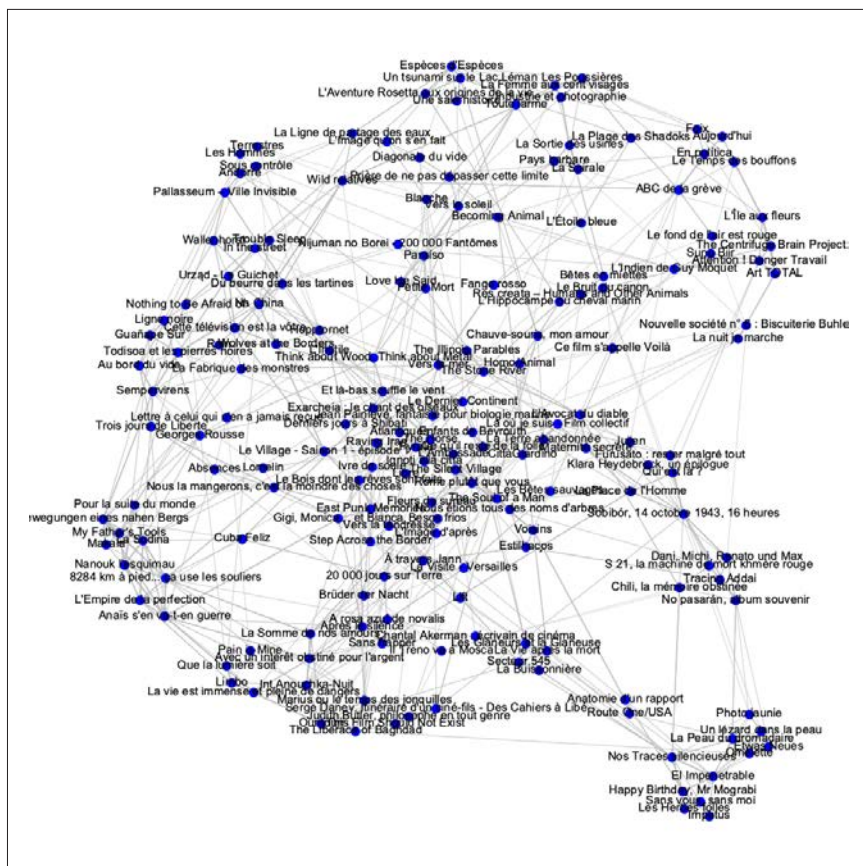
Au regard de ces deux matrices, la pondération par paliers semble la plus en adéquation avec les recommandations humaines : nous y trouvons moins de liens hors des groupes préétablis, et les liens entre films du même groupe sont plus forts et plus nombreux.

Outre les matrices d'adjacence, une autre piste explorée a été de filtrer l'affichage des liens sur un graphique nœud-lien en fonction d'un seuil de score ou d'ordre de la recommandation proposée dans Rumo. Cette solution permet d'améliorer la lisibilité des résultats de la recommandation et d'alimenter notre réflexion sur les modalités de tri des liens selon leur pertinence.

Grâce à ce tri des liens (voir la figure 10.8 à la page suivante), les groupes de films composés par la recommandation ainsi que leurs relations internes et externes ressortent plus clairement. La mise en œuvre d'un calcul de modularité à des fins de classification des films en fonction de leurs relations permet encore d'améliorer l'appréhension de ces groupes.

On retrouve dans le prochain graphique (voir la figure 10.9) la même disposition des films dans l'espace, mais le tri des liens et la classification permettent de mieux saisir le positionnement de chaque film. Les cinq films que l'on avait retenus pour leur tonalité comique se trouvent par exemple placés dans le groupe couleur cyan, en périphérie de la figure. Au contraire des films des groupes orange, rouge et violet, ils ne semblent pas faire de lien transclasses. On peut y voir la confirmation de la pertinence de la recommandation, conforme à la recommandation éditoriale. Mais cela peut être aussi interprété comme une limite de la démarche : si les films sont intégrés à des groupes de recommandation trop autonomes, l'outil ne décourage-t-il pas la découverte de nouveaux univers ?

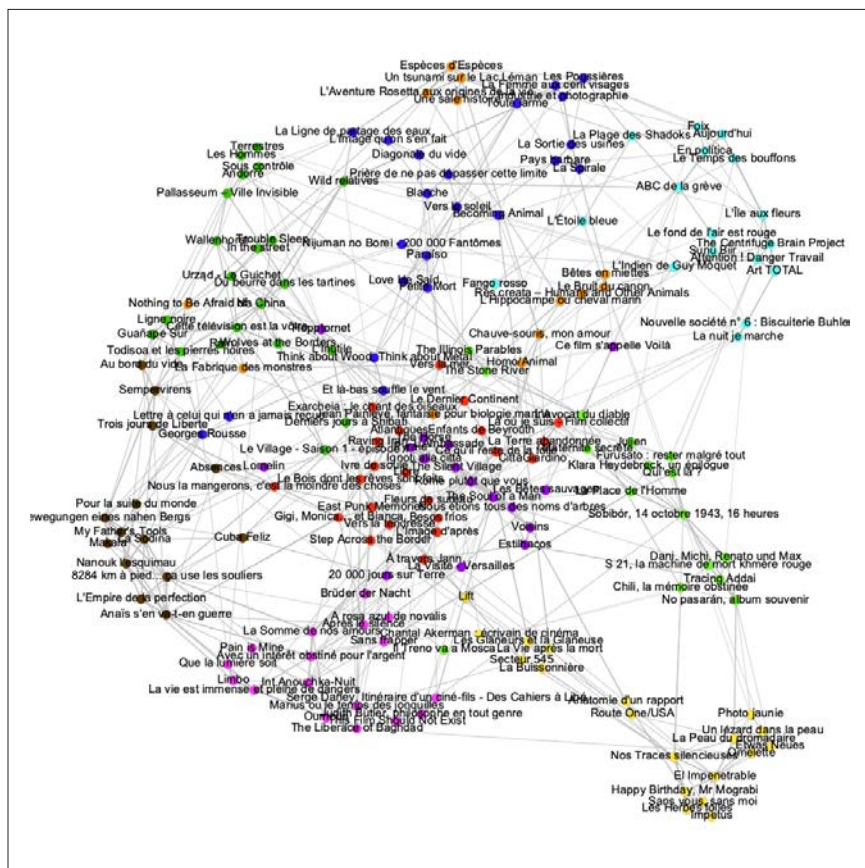
FIGURE 10.8
Représentation en réseau des liens algorithmiques
les plus forts entre films du corpus



Légende: Ce graphique représente les liens forts entre films du corpus proposés par l'algorithme de Rumo à partir de l'indexation des dispositifs cinématographiques, pondérés selon une échelle par paliers (1, 3, 5, 8, 10). La force des liens représentés est caractérisée par leur ordre, inférieur à 5, et leur score, inférieur à 0,5.

FIGURE 10.9

Représentation en réseau des liens algorithmiques les plus forts entre films du corpus, répartis par classe de modularité



Légende: Ce graphique représente les liens forts entre films du corpus proposés par l'algorithme de Rumo à partir de l'indexation des dispositifs cinématographiques, pondérés selon une échelle par paliers (1, 3, 5, 8, 10), et répartis par classe de modularité. La force des liens représentés est caractérisée par leur ordre, inférieur à 5, et leur score, inférieur à 0,5.

4. DISCUSSION : LIMITES DE LA DÉMARCHE

4.1 Une expérimentation algorithmique

Pour évaluer notre expérimentation de la recommandation automatisée de films basée sur leurs métadonnées et appréhender les résultats du calcul algorithmique, nous nous sommes trouvés face à deux difficultés.

La première semble inhérente à notre terrain : constituer un panel de prescripteurs ayant vu la totalité des films que nous avons indexés était impossible, il nous a fallu concentrer notre regard sur un petit nombre de films dont la recommandation faisait consensus auprès des spécialistes consultés.

La seconde découle de ce constat : nous avons cherché à valider les résultats de la recommandation par des outils de visualisation, notamment en réseaux, prenant ainsi le risque de superposer les couches algorithmiques de spatialisation et de classification au calcul algorithmique proposé par Rumo. Malgré cela, dès lors qu'elles sont considérées comme des outils et non comme des résultats, ces visualisations nourrissent la réflexion en donnant des pistes d'analyse.

4.2 Un algorithme trop lourd à mettre en œuvre ?

Les algorithmes de contenu à contenu se trouvent généralement écartés du fait de l'investissement humain, temporel et financier trop important à fournir pour générer des métadonnées pertinentes, et du fait de la faiblesse des métadonnées existantes (Foulonneau, Grouès, Naudet et Chevalier, 2014). Ces deux aspects peuvent être compensés par la prise en compte des traces d'usage et par la délégation aux utilisateurs de la fabrication de ces métadonnées par la folksonomie, dispositif proposant aux utilisateurs d'indexer eux-mêmes les contenus, librement ou au moyen d'une liste de termes préétablie. Toutefois, ces méthodes trouvent souvent leurs limites dans le manque de cohérence des métadonnées qu'elles produisent.

4.3 La recommandation de contenu à contenu, une technologie de l'intelligence artificielle ?

Une autre question soulevée par cette contribution concerne la nature du dispositif de recommandation considéré : relève-t-il vraiment d'une intelligence artificielle ? Oui, dans le sens où il s'agit d'un système déductif (Farchy et Denis, 2020, p. 21-22), pensé pour reproduire par l'algorithmie

un processus intelligent et remplacer un travail intellectuel humain par une action automatisée. Mais, pour mener à bien cette automatisation, un travail de documentation reste indispensable, travail dont dépend la pertinence de la recommandation algorithmique. En rejetant l'utilisation de traces d'usages, en n'automatisant pas l'indexation, en ne mobilisant pas d'algorithme prédictif ou de technologie d'apprentissage par la machine, la recommandation algorithmique de contenu à contenu dont il est question dans ce chapitre paraît finalement moins artificielle qu'humaine.

Cependant, l'automatisation de la recommandation de biens culturels au moyen d'algorithmes ne peut pas non plus être considérée comme anodine. Si les effets sociaux de tels dispositifs sont moins évidents que ceux des algorithmes utilisés pour automatiser les actions de police ou de justice, ils n'en sont pas moins réels. En donnant de la visibilité ou en n'affichant pas certains films, ils contribuent notamment à forger ou entretenir les goûts des utilisateurs des plateformes (Lawrence, 2015). Certes, ce phénomène est particulièrement sensible dans le cas de recommandations basées sur les traces d'usage (Micheau *et al.*, 2017 ; Beuscart *et al.*, 2019 ; Roberge *et al.*, 2019). Chercher à recommander des contenus à partir de leur similarité avec un autre tend toutefois également à renforcer l'exposition des usagers à un ensemble homogène de documents, et ainsi « enfermer les individus dans l'espace des possibles » (Coulangeon, 2020) dessiné par leur choix de départ.

4.4 Une multiplicité d'acteurs qui participent à la valorisation d'œuvres patrimoniales

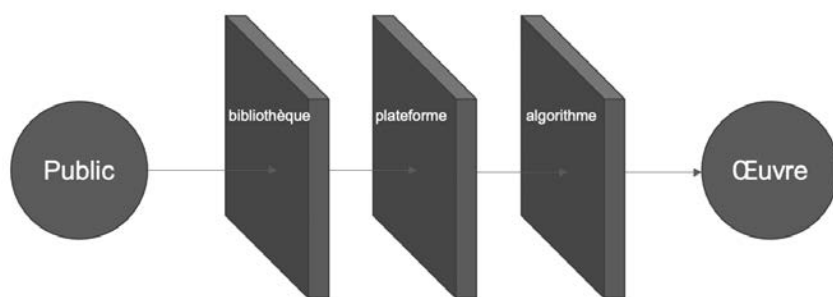
Si l'utilisation d'algorithmes de recommandation par des plateformes VAD/VADA peut servir à améliorer la découvrabilité du patrimoine cinématographique, elle peut aussi s'avérer encombrante pour les acteurs de la médiation de ces œuvres qui ne l'ont pas entamée. Dès lors que la médiation algorithmique devient incontournable pour avoir accès à l'œuvre, elle se superpose avec les autres formes de médiation, notamment humaine.

De fait, le recours à un dispositif artificiel de recommandation ne vient pas régler la question de la multiplicité des intérêts des acteurs qui participent à la valorisation d'œuvres audiovisuelles patrimoniales, aussi variés que ceux de leur potentiel public. Parmi ces intermédiaires du patrimoine cinématographique, les médiathécaires et les bibliothécaires notamment. Ils interviennent théoriquement au terme de la circulation commerciale des films, en lien direct avec leurs usagers. En sélectionnant

un film sur support physique pour l'intégrer à leurs collections, ils héritent bien sûr des actions de médiation antérieures mises en œuvre par leurs producteurs, distributeurs et éditeurs. Mais, depuis le développement de plateformes VAD/VADA, ces actions continuent de prendre place au sein même des médiathèques. Conservateurs et bibliothécaires n'y sont plus les seuls à assurer le lien entre des œuvres et leur public. Ils sont concurrencés sur ce terrain par les plateformes, qui font varier leur catalogue suivant une logique qui leur est propre, et par les dispositifs de recommandation algorithmique implantés en leur sein, qui orientent à leur tour la prescription.

FIGURE 10.10

Enchâssement des acteurs et outils de la médiation du cinéma documentaire en bibliothèques



Or la principale mission des bibliothécaires et des médiathécaires reste d'établir des liens entre usagers et œuvres, voire entre usagers :

On ne peut pas faire l'économie de ce lien avec l'utilisateur. Lorsqu'on pense un accès à distance comme la VOD, il faut réfléchir à quel lien on maintient avec les usagers. On ne peut pas simplement rester à un lien qui serait celui d'un algorithme¹⁷.

Dans cette perspective, le recours à des algorithmes de recommandation par les plateformes semble un frein à l'activité de médiation des professionnels de bibliothèques. Pour autant, les dispositifs de recommandation

17. Jean-Yves de Lepinay (président d'Images en bibliothèque), « Quelle place et quelles missions pour les bibliothèques ? Appréhender les différentes offres et l'évolution des pratiques en médiathèque », *VàD et ressources numériques*, formation professionnelle organisée en ligne par l'association Images en bibliothèque à destination de médiathécaires, 29 avril 2021.

algorithmiques pourraient également être mobilisés directement par ces professionnels pour les aider dans leurs missions, en particulier pour suggérer de nouveaux liens entre films, au même titre que les outils documentaires qu'ils utilisent au quotidien.

CONCLUSION

L'efficacité des algorithmes de contenu à contenu pour faciliter la médiation du patrimoine cinématographique est dépendante des données auxquelles il est appliqué. Les liens entre films qu'ils mettent en avant sont avant tout tissés par les stratégies documentaires établies en amont par les structures qui les implémentent. Ils permettent toutefois, combinés à des outils de visualisation de la recommandation, d'adapter la documentation sciemment. Ils constituent donc une option plus éthique et plus humaine que les algorithmes de recommandation classiquement basés sur les traces d'usage d'utilisateurs. En perpétuant des usages sociaux lui préexistant, le recours à un algorithme de recommandation de contenu à contenu ne résout pas la difficulté à concilier les logiques concurrentes des acteurs de la médiation du cinéma patrimonial. Mais il constitue un outil comme un autre dont les acteurs de la médiation traditionnelle peuvent aussi s'emparer pour en faire une technologie collective conforme à leurs valeurs.

RÉFÉRENCES

- Adary, Assaël et Jean-Claude Domenget, « Entretien : pour une intégration des datadéontologues dans les équipes communication », *Revue Communication & professionnalisation*, n° 6, 2017, p. 152-160.
- Bachimont, Bruno, « Nouvelles tendances applicatives : de l'indexation à l'éditorialisation », dans Patrick Gros (dir.), *L'indexation multimédia : description et recherche automatiques*, Paris, Lavoisier, 2007, p. 313-326.
- Benghozi, Pierre-Jean et Thomas Paris, « L'économie culturelle à l'heure du numérique : une révolution de l'intermédiation », dans Laurent Jeanpierre et Olivier Roueff (dir.), *La culture et ses intermédiaires : dans les arts, le numérique et les industries créatives*, Paris, Éditions des archives contemporaines, 2014, p. 175-188.
- Benghozi, Pierre-Jean et Thomas Paris, « The cultural economy in the digital age : A revolution in intermediation ? », *City, Culture and Society*, vol. 7, n° 2, 2016, p. 75-80.
- Beuscart, Jean-Samuel, Samuel Coavoux et Sisley Maillard, « Les algorithmes de recommandation musicale et l'autonomie de l'auditeur », *Réseaux*, n° 213, 2019, p. 17-47.

- Bonenfant, Maude, Lucile Crémier et Laura Iseut Lafrance St-Martin, « Quelques réflexions sémiotiques sur le circuit des données massives », dans André Mondoux et Marc Ménard (dir.), *Big Data et société : industrialisation des médiations symboliques*, Québec, Presses de l'Université du Québec, 2018, p. 131-163.
- Bullich, Vincent, « L'intégration de la production cinématographique aux technologies numériques de l'information et de la communication », dans Michaël Bourgatte et Vincent Thabourey (dir.), *Le cinéma à l'heure du numérique : pratiques et publics*, Paris, MkF Éditions, 2012, p. 23-43.
- Bullich, Vincent et Viviane Clavier, « Production des données, « Production de la société ». Les Big Data et algorithmes au regard des Sciences de l'information et de la communication », *Les enjeux de l'information et de la communication*, n° 19/2, 2018, p. 5-14.
- Cardon, Dominique, *À quoi rêvent les algorithmes : nos vies à l'heure des big data*, Paris, Seuil, 2015.
- Casilli, Antonio A., *En attendant les robots : enquête sur le travail du clic*, Paris, Seuil, 2019.
- Coulangeon, Philippe, « Comment se forment les goûts culturels aujourd'hui ? » dans Serge Paugam (dir.), *50 questions de sociologie*, Paris, Presses universitaires de France, 2020.
- Danard, Benoît, *L'Économie des films de patrimoine*, Paris, France, CNC, 2020.
- Delcroix, Geoffrey, « Les données et les algorithmes, nouvelles muses de la prescription culturelle », dans Brigitte Chapelain et Sylvie Ducas (dir.), *Prescription culturelle : avatars et médiamorphoses*, Villeurbanne, ENSSIB, 2018, p. 347-359.
- Dessinges, Catherine et Lucien Perticoz, « Les consommations de séries télévisées des publics étudiants face à Netflix : une autonomie en question », *Les enjeux de l'information et de la communication*, n° 20/1, 2019, p. 5-23.
- Direction des études, des statistiques et de la prospective, *Le bilan 2020 du CNC*, Paris, CNC, 2021.
- Doueïhi, Milad, *Qu'est-ce que le numérique ?*, Paris, Presses universitaires de France, 2013.
- Du Val, Jean-Luc, « "Toutes les images du monde" au plus près de chez soi : donner accès aux films dans les bibliothèques », dans Dominique Rousselet, Julie Guillaumot et Marianne Palesse (dir.), *Du cinéma en bibliothèque*, Paris, ABF et Images en bibliothèques, 2017, p. 69-76.
- Epron, Benoît et Marcello Vitali-Rosati, « Pour une définition de l'édition à l'ère numérique », dans *L'édition à l'ère numérique*, Paris, La Découverte, 2018, p. 5-33.
- Farchy, Joëlle et Juliette Denis, *La culture des données : intelligence artificielle et algorithmes dans les industries culturelles*, Paris, Presses de l'École des mines, 2020.
- Flichy, Patrice, « Produire les données, à la frontière de plusieurs mondes sociaux », *Réseaux*, n° 228, 2021, p. 9-19.
- Foulonneau, Muriel, Valentin Grouès, Yannick Naudet et Max Chevalier, « Recommandeurs et diversité : exploitation de la longue traîne et diversité des listes de recommandation », dans Gérald Kembellec, Ghislaine Chartron et Imad Saleh (dir.), *Les moteurs et systèmes de recommandation*, London, ISTE Éditions, 2014, p. 85-104.
- Gantier, Samuel, « Construction de pratiques cinéphiles sur une plateforme de vidéo à la demande : enjeu du design des personae », *Les enjeux de l'information et de la communication*, vol. 21, n° 1, 2020, p. 53-73.

- Gantier, Samuel, Ève Givois et Éric Kergosien, « Recommander autrement les films documentaires? Design d'une preuve de concept basée sur l'indexation des dispositifs de réalisation », *RIHM*, vol. 21, n° 2, 2020, p. 1-26.
- Givois, Ève, Laure Bolka-Tabary, Éric Kergosien, Samuel Gantier et Bouchra El Haoudi, « Élaboration itérative d'un thésaurus pour indexer le cinéma documentaire à partir des procédés filmiques », *I2D*, n° 2, 2021, p. 122-150.
- Henry, Nathalie et Jean-Daniel Fekete, « Représentations visuelles alternatives pour les réseaux sociaux », *Réseaux*, vol. 26, n° 152, 2008, p. 59-92.
- Jeanpierre, Laurent et Olivier Roueff (dir.), *La culture et ses intermédiaires : dans les arts, le numérique et les industries créatives*, Paris, Éditions des archives contemporaines, 2014.
- Kilborne, Yann, *L'expérience documentaire : approche communicationnelle du cinéma de réalité*, thèse de doctorat en sciences de l'information et de la communication dirigée par Jean-François Dégremont, Université Paris 8 Saint-Denis, 2008.
- Laurent, Natacha, « Le patrimoine est-il soluble dans le marché? », dans Christophe Gauthier et al. (dir.), *Patrimoine et patrimonialisation du cinéma*, Paris, École nationale des chartes, 2020a.
- Laurent, Natacha, « "L'ivresse patrimoniale" et après? », 1895, vol. 91, n° 2, 2020b, p. 8-30.
- Lawrence, Emily, « Everything is a recommendation. Netflix, altgenres and the construction of taste », *Knowledge Organization*, vol. 42, n° 5, 2015, p. 358-64.
- Le Deuff, Olivier, « Folksonomies : les usagers indexent le web », *Bulletin des bibliothèques de France*, n° 4, 2006, p. 66-70.
- Le Ludec, Clément, Elinor Wahal, Antonio A. Casilli et Paola Tubaro, « Quel statut pour les petits doigts de l'intelligence artificielle? Présent et perspectives du micro-travail en France », *Les Mondes du travail*, n°s 24-25, 2020, p. 99-112.
- Lesaunier, Marie-Eva, « Une plateforme au service d'un monde professionnel mobilisé : enjeux symboliques et économiques », *Tic & société*, vol. 13, n°s 1-2, 2019, p. 225-246.
- Lindeperg, Sylvie et Ania Szczepanska, *À qui appartiennent les images? Le paradoxe des archives, entre marchandisation, libre circulation et respect des œuvres*, Paris, Les Éditions de la MSH, 2017.
- Menger, Pierre-Michel, « Introduction », dans Simon Paye (dir.), *Big data et traçabilité numérique. Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France, 2017, p. 7-23.
- Micheau, Béatrice, Marie Després-Lonnet et Dominique Cotte, « La recommandation musicale entre inscriptions documentaires, pratiques sociales et dispositifs d'écoute », *Études de communication : langages, information, médiations*, n° 49, 2017, p. 33-56.
- Pariser, Eli, *The Filter Bubble : How the New Personalized Web Is Changing What We Read and How We Think*, London, Penguin Books, 2011.
- Roberge, Jonathan, Romuald Jamet et Andréanne Rousseau, *L'impact social des algorithmes de recommandation sur la curation des contenus musicaux francophones au Québec : enquête qualitative*, Québec, Institut national de la recherche scientifique, 2019.
- Simonnot, Brigitte, *Conduire des recherches en régime numérique : vers un cadre conceptuel de réflexion éthique*, Bruxelles, De Boeck Supérieur, 2018.

- Taillibert, Christel, « Festivals et patrimoine cinématographique à l'heure du numérique. La situation française », *Secuencias. Revista de historia del cine*, n° 39, 2015, p. 83-99, traduction française de l'autrice, <<https://hal.archives-ouvertes.fr/halshs-03161036/>> (consulté le 9 février 2022).
- Taillibert, Christel, *Vidéo à la demande : une nouvelle médiation ? Réflexions autour des plateformes cinéphiles française*, Paris, L'Harmattan, 2020.
- Taillibert, Christel, « Penser l'accompagnement éducatif cinéphile dans un environnement numérique », *Communiquer. Revue de communication sociale et publique*, n° 31, 2021, p. 75-96.
- Thuillas, Olivier et Louis Wiart, « Les plateformes de VOD cinéphiliques. Des stratégies de niche en questions », *Les Enjeux de l'information et de la communication*, n° 20/1, 2019a, p. 39-55.
- Thuillas, Olivier et Louis Wiart, « Plateformes alternatives et coopération d'acteurs : quels modèles d'accès aux contenus culturels ? », *Tic & société*, vol. 13, n°s 1-2, 2019b, p. 13-41.
- Tiffon, Guillaume, « La contribution des internautes aux big data : un travail ? », dans Pierre-Michel Menger et Simon Paye (dir.), *Big data et traçabilité numérique : les sciences sociales face à la quantification massive des individus*, Paris, Collège de France, 2017, p. 69-81.
- Vander Wal, Thomas, « You down with folksonomy ? », 4 août 2004, <<https://www.vanderwal.net/random/entrysel.php?blog=1529>> (consulté le 10 février 2022).
- Vayre, Jean-Sébastien, « Une histoire de machines à recommander des biens de consommation. De l'Internet documentaire à l'Internet des données », *Études de communication : langages, information, médiations*, n° 49, 2017, p. 89-106.
- Vayre, Jean-Sébastien, « Comment décrire les technologies d'apprentissage artificiel ? », *Réseaux*, n° 211, 2018, p. 69-104.

LES PARTENARIATS DE DONNÉES, UNE PISTE DE SOLUTION COLLECTIVE AUX ENJEUX DE DÉVELOPPEMENT DES APPLICATIONS DE L'IA EN ART ET EN CULTURE

Viêt Cao¹ et Juliette Denis²

1. PROBLÉMATIQUE

L'intelligence artificielle (IA) a de nombreuses applications en arts et en culture. Elle repose toutefois grandement sur les données dont elle se nourrit, en particulier en matière de quantité de données dont les organismes culturels de petite ou moyenne taille ne disposent pas à une échelle individuelle. Pour tenir ses promesses de performance et de pertinence, des générations de données de qualité interoperables sont nécessaires, sans compter les risques de biais de représentation et de modélisation qui en découlent. Parmi les pistes de solutions, les partenariats de données entre acteurs culturels et avec la communauté de développement de l'IA constituent une solution collective à co-développer, intégrer et mettre en œuvre.

1. Directeur innovation et analytique de données de Synapse C.
2. Gestionnaire de projets collaboratifs chez Synapse C.

2. L'INTELLIGENCE ARTIFICIELLE EN ARTS ET EN CULTURE

L'IA est un outil, un objet et un sujet pour l'art et la culture. Technologie disruptive, l'IA permet la création d'images et de peintures, de sons et de musiques, la génération de mots et de textes, et ces créations de l'IA révèlent ce qui sépare ou rapproche l'homme de l'IA. Parce qu'elle pose les questions de droits d'auteur, de rémunération de l'IA sur ces créations, ou encore la valeur du travail humain, l'IA est un objet et un sujet de choix à examiner pour les artistes. En créant les espaces d'utilisation, d'expérimentation et de débat autour de l'IA, la communauté artistique participe déjà aux développements de l'IA. Si l'IA est le reflet de la société humaine, elle est aussi attendue dans sa capacité à ne pas en reproduire certains biais (conscients ou inconscients).

Les bénéfices tangibles à court terme de l'IA sont de pouvoir apprendre des tendances et des corrélations dans une variété de contextes à partir de données ou d'automatiser certaines opérations répétitives. Mis à part les activités de création mentionnées ci-dessus, les bénéfices recherchés sont pour la plupart des acteurs des industries culturelles et créatives d'agrandir le bassin d'utilisateurs ou d'atteindre de nouvelles cibles ou de nouveaux segments ou groupes (auditeurs, lecteurs, consommateurs) dans un contexte changeant en permanence : de nouvelles œuvres sont créées tous les jours, les goûts des publics évoluent constamment. Pour les utilisateurs, l'IA aide à naviguer et à découvrir des œuvres et des contenus, dans un contexte de mondialisation et d'abondance de contenus culturels physiques et numériques. Aujourd'hui, l'IA influence les choix des consommateurs, ce qui influe sur la création artistique par effet de levier économique.

3. LES TROIS ÉLÉMENTS D'UNE SOLUTION D'IA

Une solution d'IA inclut trois éléments : les données, l'algorithme (ou code) et l'exécution (déploiement). Si la prévalence des données ou de l'algorithme dans la solution d'IA n'est pas clairement quantifiable (Digital Curation Center, 2020), les données en demeurent un élément central. C'est pourquoi leur préparation et leur compréhension sont importantes et améliorent les performances de la solution. Il a été documenté que ce n'est pas tant la quantité de données que d'avoir la bonne donnée et des données de qualité qui importe (Ghosh, 2019 ; Redman, 2018) : « *garbage in, garbage out* ». Choisir une solution d'IA, c'est choisir le modèle qui a été développé avec un jeu de données précis et particulier.

3.1 Le défi d'intégration des données

Réunir ces trois éléments n'est pas trivial, les équilibrer l'est encore moins. En ce qui concerne les données, un défi d'opérationnalisation pour la plupart des organismes est d'appliquer la solution d'IA aux besoins d'affaires réels, avec des données réelles. Une condition de réalisation de la promesse de « performance » de l'IA est en effet la nécessité de grands ensembles de données pertinentes, si bien que seulement peu d'entreprises verticales nées avec le numérique ont réussi à la mettre en place (ex. : Amazon, Hulu, Netflix, Spotify, Pandora, YouTube). Bien que l'IA moins intensive en données se développe (*data-hungry*), elle est encore loin d'être prête à l'opérationnalisation. Ainsi, au lieu d'avoir des relations multivaleurs (*many-to-many*) entre les œuvres et les audiences (une œuvre pour chaque public, un public pour chaque œuvre), on se retrouve plutôt dans des relations un à plusieurs (*one-to-many* : tous les auditoires sont dirigés vers une minorité d'œuvres) ou des rapports de plusieurs à un (*many-to-one* : les œuvres sont créées et diffusées pour un certain type de public et non pour plusieurs types de publics).

Or, la curation de la culture est la résultante de compromis multiples entre les créateurs, les diffuseurs/distributeurs, ainsi que les usagers (Born *et al.*, 2021). Le rôle en culture que porte l'opérateur de la solution d'IA va fortement influencer sur les objectifs de l'algorithme, son déploiement et son évaluation. Par exemple, un créateur émergent peut vouloir augmenter sa notoriété et non ses ventes, tandis qu'une plateforme peut vouloir agrandir sa base d'utilisateurs, capter l'attention, optimiser des revenus publicitaires, tandis qu'un usager va préférer une rapidité et une pertinence de recommandation.

Ainsi, on retrouve dans les industries culturelles tous les grands défis de développement de l'IA :

- Explicabilité et transparence des solutions d'IA (objectifs, décisions ou médiation) ;
- Responsabilité et éthique ;
- Sobriété en données (accessibilité pour des organismes de toutes tailles ou moyens, minimiser les collectes de données) ;
- Meilleurs algorithmes.

3.2 Les enjeux de la concentration des acteurs et des données

Dans l'économie numérique, la question de l'accès aux données et des asymétries critiques entre les acteurs se révèle déjà cruciale («les données sont le nouveau pétrole»). Que ce soient les mêmes acteurs qui concentrent les données décrivant les contenus culturels, les données décrivant les usagers tout en développant les algorithmes et les solutions d'IA, est un danger réel. Les types de biais inhérents aux modèles d'IA et l'intérêt privé influent sur la diversité des contenus, leur médiation vers les usagers et leur découvrabilité locale et globale.

Suresh et Gutttag (2019) décrivent six types de biais dans les solutions d'IA et leur déploiement : on retrouve, en ce qui concerne les données, les biais historiques, les biais de représentation et les biais de mesure qui représentent «la» réalité ; au niveau de l'algorithme, les biais d'agrégation et les biais d'évaluation ; tandis qu'en ce qui concerne l'exécution, on retrouve les biais de déploiement de la solution d'IA.

En vue d'examiner ces enjeux, ce n'est pas une entité unique et intégrée, mais de multiples entités qui doivent être parties prenantes à tous ces niveaux de développement de solutions d'IA appliquées en art et en culture. Dans ce texte, on désignera par simplification «organismes culturels» les entreprises, les départements ou les unités en lien avec les activités de création, médiation et diffusion de produits ou contenus culturels et par «entité IA» les entreprises, fournisseurs technologiques ou départements qui touchent directement à l'algorithme et l'exécution de la solution d'IA. Dans cette typologie, Spotify est ainsi une entité IA, car elle ne crée ou ne produit pas de chansons, tandis que Netflix est à la fois un organisme culturel (rôle de producteur) et une entité IA (rôle de service de diffusion).

4. LES PARTENARIATS DE DONNÉES, UNE PISTE DE SOLUTION À LA PORTÉE DE TOUS

Les organismes culturels ne vont probablement pas maîtriser toutes les techniques et les infrastructures nécessaires au développement d'algorithmes et à leur déploiement et pourraient donc avoir plus de maîtrise et de contrôle sur la génération, la collecte, le traitement et l'interprétation des données. L'éthique de l'utilisation des données devient

un sujet réservé aux acteurs techniques; l'UNESCO a dédié une section dans sa *Recommandation sur l'éthique de l'intelligence artificielle*³.

La quantité et la qualité des données mobilisables pour alimenter les algorithmes sont en effet deux facteurs essentiels de la qualité et de la compétitivité de la solution développée, tandis que la capacité à réunir les parties prenantes de leur utilisation et leur déploiement devient une condition essentielle à des effets équitables et éthiques. Ainsi, les « partenariats de données » constituent une piste de solution :

- en multipliant les parties prenantes et les intervenants autour d'une donnée (participation à la collecte de données, leur annotation, leur interprétation, leur modélisation et leur utilisation), autour des entités IA et autour de la solution d'IA à déployer ;
- en s'appuyant sur les ensembles de données réelles (en général de tailles modestes) à combiner pour proposer de nouveaux ensembles de données avec moins de biais de représentation et répondant aux exigences des modèles d'IA et aux objectifs d'affaires.

4.1 Nécessité d'appropriation et d'annotation des données par le secteur culturel

Le biais de représentation et le biais de mesure sont des enjeux réels et d'autant plus probables lorsque les données représentant les œuvres ne sont pas créées par les acteurs du milieu eux-mêmes. Les organismes culturels, producteurs de contenus, doivent participer, voire mener les travaux de représentation numérique des œuvres créées – par exemple : la communauté LODEPA pour décrire les événements culturels sur le Web (Estermann, 2020) ou encore la Banque de titres de langue française (BTLF) pour les livres francophones au Québec. À l'instar de la production des données, l'offre et le patrimoine culturel s'enrichissent de jour en jour et il convient d'orienter les intermédiaires et les usagers avec des contenus repérables et accessibles. Les solutions d'IA permettent de répondre à cet enjeu, mais il convient d'abord d'orienter les créateurs de ces solutions afin de les amener aux publics finaux. Les rôles d'experts de domaine (*subject matter expert*), ou de traducteur de données (*data translator, analytics translator, data strategist*) émergent de plus en plus au sein des entreprises pour augmenter l'effet et la pertinence des modèles d'IA développés par

3. UNESCO, *Recommandation sur l'éthique de l'intelligence artificielle*, SHS/BIO/PI/2021/1, 23 novembre 2021.

rapport au contexte d'affaires auquel ils sont destinés (Henke *et al.*, 2018). La culture ne fait pas exception.

Aujourd'hui, l'IA intervient à toutes les étapes de la chaîne de valeur en arts et en culture (création, production, distribution, consommation). La technologie est en effet disruptive et crée des possibilités à faible coût d'entrée : un musicien peut de nos jours s'autoproduire ou un auteur s'aut éditer, sans passer respectivement par des studios professionnels ou un éditeur traditionnel. En internalisant ces expertises humaines, la technologie les met par ailleurs en danger. L'automatisation en IA peut en effet réduire cette dépendance aux experts (Caramiaux, 2020).

4.2 Des données interopérables s'enrichissent mutuellement

Les données descriptives des œuvres générées ne peuvent demeurer en silo : elles sont mobilisées par d'autres systèmes et acteurs. Ainsi, elles doivent de plus circuler de manière fluide en raison des multiples systèmes concernés et pour mettre à profit les espaces et les systèmes numériques du Web (cf. données ouvertes et liées, réseaux sociaux, sites Internet). Cela requiert des organismes culturels un apprentissage afin de générer des données en langage machine, standardisées pour les contenus actuels, futurs, mais également passés.

Dans un contexte de mondialisation, le numérique est une occasion de développement des publics au-delà du local, toutefois avec une concurrence accrue car il est planétaire. Les données descriptives des contenus (qui décrivent une œuvre) doivent être de qualité égale pour le plus d'œuvres possible pour offrir une chance équitable à leur curation pour les algorithmes d'IA.

Les données d'usage qui décrivent les comportements des consommateurs générés par le déploiement de ces solutions d'IA (ex. : nombre d'écoutes d'une chanson, commentaires ou mentions « J'aime » d'un livre, type d'utilisateur, etc.) ne devraient pas tendre à la simplification ou à la sous-catégorisation des usagers, mais à en étendre la richesse afin de mieux comprendre différents segments d'usagers et de les servir de manière équitable et inclusive.

Enfin, les données décrivant l'interaction entre l'œuvre et l'utilisateur gagneraient à être mieux décrites, dans la mesure où elles sont infinies (aspect d'intérêt, esthétique, aspects commerciaux, appréciation selon l'historique de chacun, contexte...). Ainsi, une chanson n'est pas qu'une

suite de fréquences et de rythmes, les paroles peuvent avoir une signification différente d'un utilisateur à un autre (valeur sentimentale). Par exemple, Pandora utilise le Music Genome Project pour décrire les chansons à partir de descripteurs maison appelés « gènes », mais leur description est tenue secrète tandis que Nielsen les décrit différemment avec Gracenote.

De meilleures données générées par les acteurs pertinents permettraient de diminuer les biais (biais historique, biais de représentation, biais de mesure au niveau des données qui représentent « la » réalité, les biais d'agrégation), d'améliorer la médiation telle que voulue par les créateurs, et non déformée par les intermédiaires/médiateurs, de participer à améliorer l'interprétation des résultats et d'orienter le développement de modèles sur les « bonnes » données (et non l'inverse qui consiste à trouver ou à générer des données qui correspondent aux besoins de l'algorithme).

4.3 La gouvernance et la gestion de données au cœur des partenariats de données

Mettre en application ces pistes d'actions implique des enjeux de génération, collecte, accès, partage et audit des données du secteur culturel. Plusieurs voies peuvent être explorées, selon les contextes commerciaux, légaux, voire éthiques. Toutes mobilisent des notions de gouvernance et de gestion de données existantes. Les contextes peuvent en effet varier selon :

- Les types de données : les données privées, les données partagées, les données mutualisées, les données ouvertes ;
- Les configurations de partenariats de données : le collectif de données, la coopérative de données, la fiducie de données, les communs numériques.

Les modèles principaux de partenariat de données en émergence sont présentés, ainsi que des facteurs critiques à leur opérationnalisation. La mise en œuvre d'un partenariat de données, notamment la gouvernance des données, touche en effet à plusieurs aspects : les objectifs du partenariat de données, les parties prenantes qui gèrent les données, la confidentialité et la sécurité des données, la propriété intellectuelle des bases de données et des applications produites avec ces données.

5. PRÉSENTATION DE QUELQUES MODÈLES DE PARTENARIATS DE DONNÉES

Le *collectif de données* est une forme de partenariat regroupant des entreprises privées, des établissements de recherche et publics, visant à

combiner les données pour générer de la valeur publique (orientée sur un objectif). La *coopérative de données* est un groupe de personnes se rassemblant pour atteindre des objectifs communs, avec une organisation. Ce sont des mutuelles détenues et exploitées par les membres (orientées sur le contrôle des données et des objectifs). La *fiducie de données* est une structure juridique pour assurer une intendance indépendante des données, au service et au bénéfice des fiduciaires (orientée sur la transparence et la confiance). Les *communs numériques* sont des initiatives au sein desquelles les données sont partagées comme ressources communes (orientées sur le cadre et le contexte de partage).

Ces modèles de partenariats de données mettent l'accent sur un aspect ou un autre (données, objectifs), avec plus ou moins de contrôle et plus ou moins de définition dans l'objectif ou l'utilité désirés. Ils peuvent donc mieux répondre à une idée, une phase de projet, un degré de maturité ou un objectif d'affaires. Ils sont décrits plus exhaustivement par Gagnon-Turcotte et ses collaborateurs (2020).

5.1 Une tâche et une responsabilité collective

Une pluralité et une diversité des sources de données sont nécessaires pour alimenter et entraîner les modèles d'IA avec plus de transparence, voire de qualité et de pertinence. Cela repose sur une disponibilité et une interopérabilité (sémantique, technique) des données afin d'agrandir l'espace d'entraînement, mais surtout sur des configurations multi-organismes afin de diversifier les sous-espaces le constituant. Pour ce faire, un dialogue local et global des acteurs doit se créer, pour des bénéfices locaux et mondiaux.

Ci-dessous, trois exemples de dérives et dangers de l'application de l'IA en culture sont présentés, de même que des propositions de solutions au moyen de partenariats de données.

CAS 1 : ALGORITHME DE RECOMMANDATION

Un exemple commun en arts et culture est l'algorithme de recommandation, qui est un intermédiaire algorithmique entre des œuvres et des usagers. Différentes logiques peuvent sous-tendre leur développement et leur fonctionnement : un algorithme centré sur la plateforme (objectif de garder captif, de maximiser l'usage), un algorithme centré sur le contenu (objectif que le contenu « trouve » une audience) ou un algorithme centré sur l'utilisateur (objectif d'avoir une réponse

rapide en tout temps, potentiellement au détriment de la pertinence). Parfois, tous ces objectifs d'affaires coexistent et leur pondération dans l'algorithme est opaque (boîte noire). L'évaluation de la performance et de la pertinence est directement en lien avec la logique d'affaires, ce qui peut entraîner des biais.

Par exemple, Spotify est assez généraliste en musique et n'est pas centré uniquement sur la plateforme, le contenu ou l'utilisateur. Il mise sur l'exploitation de la connaissance de l'utilisateur (historique, chansons non écoutées, pays de résidence) et l'exploration de nouvelles propositions musicales à partir du comportement des autres utilisateurs pour offrir une expérience personnalisée (Born *et al.*, 2021). Pandora semble plus axé sur la découverte de musique pour les utilisateurs, en se fondant sur plus de 450 descripteurs de chansons, développés avec des musicologues (Latour, 1999). Dans les deux cas, les recommandations sont limitées structurellement par l'offre du catalogue et par les usagers abonnés au service (Born *et al.*, 2021).

Des partenariats de données pourraient être mis en place avec les établissements de recherche afin de tester les algorithmes sur de multiples catalogues. Par exemple au Québec, le collectif de données MétaMusique s'est entendu avec l'industrie musicale pour développer un référentiel commun de données descriptives pour la musique (MétaMusique, 2021). Ces données, produites avec une certaine qualité et logique, ont pour objectif d'influencer la découvrabilité de la musique québécoise. L'initiative en elle-même devra établir un modèle d'affaires afin de trouver un équilibre entre les ressources nécessaires pour produire de la donnée de qualité, pour des bénéficiaires tiers.

Le projet TAMIS dans le domaine du livre vise à créer des données descriptives sur les œuvres littéraires, plus centrées sur l'œuvre que sa forme ou son support, et les générer en grand volume, avec une forme d'automatisation (Tamis, 2022). L'IA est ici utilisée pour générer des données, qui sont à l'étude pour être diffusées en données ouvertes et liées par Wikidata.

CAS 2 : DIVERSITÉ DES CONTENUS ET DES BULLES DE FILTRE

Les créateurs ont potentiellement accès à tous les usagers et tous les publics, mais l'inverse n'est pas toujours vrai. En ce sens, tout intermédiaire est un médiateur et un curateur culturel. Si les plateformes et services de diffusion constituent des vitrines qui augmentent la visibilité et la découvrabilité, les algorithmes réduisent la diversité en présentant de manière personnalisée et ordonnancée un nombre limité d'œuvres recommandées selon des critères propres à la logique encodée dans l'algorithme. Il devient crucial que ces algorithmes ne fonctionnent pas tous de la même façon afin de préserver une pluralité d'approches : de même qu'une seule culture ne devrait pas être hégémonique (cf. enjeux de diversité

des expressions culturelles : on peut voir une carte du monde centrée sur l'Europe ou encore sur un autre continent ou un unique grand océan entouré de continents telle que la carte de Spilhaus), une homogénéisation des façons de développer les algorithmes réduirait la diversité des contenus culturels présentés et consommés. Par exemple, l'algorithme de YouTube ne favorise pas l'ensemble des vidéos qui existent : il a été reporté que plus de 90 % des vidéos existantes sur YouTube n'avaient que 1000 vues ou moins, alors qu'il y a plus de 5 milliards de visionnements par jour (Wilde, 2020). Environ 10 000 vidéos ont au contraire atteint ou dépassé le milliard de visionnements. Ainsi, il n'est pas souhaitable que tous les algorithmes se calquent sur celui de YouTube. De plus, les approches de recommandations fondées sur la consommation et la similarité entre usagers renforcent et amplifient ces biais : on va avoir tendance à vouloir reproduire un comportement type.

Des partenariats de données pourraient être mis en place entre les développeurs de ces algorithmes afin qu'ils soient testés par différents acteurs, avec différentes œuvres ou bases d'utilisateurs, et en analyser les réponses. Au niveau scientifique, des approches d'explicabilité des modèles pourraient être ainsi testées (par exemple, les approches SHAP [2022] ou LIME [Ribeiro *et al.*, 2016]), tandis qu'au niveau des contextes d'application des recherches pourraient être approfondies. Par ailleurs, opérer un même algorithme avec différents jeux de données structurées différemment permet de diversifier son « entraînement » (approche centrée sur les données), mais également les contextes où il est déployé, et de travailler l'algorithme afin qu'il puisse apprendre de données réelles, et non d'aller chercher les données requises par l'algorithme (approche centrée sur le modèle). La Culture crée (*Culture Creates*, 2022) et le projet Avenir numérique lié de CAPACOA (2022) fédèrent plusieurs efforts afin de générer des données pour les spectacles en arts de la scène au Canada ; le modèle de coopérative de données est envisagé pour l'exploiter à terme.

CAS 3 : DES DONNÉES D'ENTRAÎNEMENTS ET DE DÉVELOPPEMENT COMMUNS

Publier en données ouvertes ou en communs numériques est une façon de participer aux développements de l'IA en fournissant des données d'entraînement ouvertes à tous. À titre d'exemples, mentionnons MovieLens pour l'industrie du film, Book-Crossings ou Goodreads pour l'industrie du livre.

Plus généralement, les données générées par les systèmes d'IA peuvent être utiles pour des organismes autres que ceux qui les exploitent. Dans une ère abondante de données, les pouvoirs publics sont paradoxalement pauvres en information et ils n'ont pas de vue d'ensemble des habitudes de consommation

(musique québécoise, ventes de livres francophones numériques, etc.). Des partenariats de données sont nécessaires afin de combler ces manques. Ainsi, aucun organisme autre qu'Alphabet/Google, qui domine les activités de recherche sur Internet, ne pourrait avoir une idée des habitudes de recherche de contenus culturels au pays. Ces données pourraient être utiles aux organismes culturels pour comprendre et développer leur non-public. Ceux-ci pourraient tenter de mettre ensemble toutes leurs données Google Analytics dans le cadre d'un partage ou d'une mutualisation des données.

Un autre bénéfice commun serait de constituer une masse critique d'entraînements de contenus en français québécois, pour ré-entraîner des modèles de traitement du langage naturel (NLP) développés dans d'autres langues, ou même en français de France.

En vue de permettre ce dialogue, les établissements publics et universitaires doivent créer des conditions pour des recherches alternatives, pour auditer et remettre en question, pour permettre l'exploration. En ce qui concerne les données, un certain soutien financier des pouvoirs publics est nécessaire. L'enjeu est économique: le plus souvent, ceux qui arrivent à tirer des bénéfices des données ne sont pas ceux qui ont généré les données. Produire des métadonnées au niveau national, au bénéfice d'acteurs privés, est coûteux, mais empêche de dépendre de ces acteurs privés pour représenter numériquement et effectuer la médiation des œuvres et des contenus.

5.2 Rôles des organismes culturels

Ainsi, par la génération de données et les partenariats de données, les organismes culturels peuvent agir en amont sur les données décrivant les œuvres et en aval sur les données décrivant leurs usagers, influant ainsi les algorithmes qui mobilisent ces données, tout en créant de nouvelles données. C'est avec un dialogue à créer et à entretenir avec les entités IA que les besoins d'affaires seront bien programmés, intégrés et évalués dans le code, tandis que le déploiement est à faire ensemble pour assurer la pertinence de la solution aux besoins d'affaires. Si l'IA est une technologie qui peut paraître très abstraite, sa mise en œuvre est encore profondément humaine (Born *et al.*, 2021). L'Union européenne a établi en 2020 un premier cadre d'auto-évaluation de la confiance ALTAI pour toute solution d'IA mise en œuvre par n'importe quelle entreprise (Altai, 2022).

Les organismes culturels doivent aussi être exigeants avec les solutions qu'ils font développer par les entités IA, afin de préserver les œuvres et les

usagers. Ils doivent être prudents avec les données des usagers. Si des idées et des technologies se développent pour être moins intensives en données d'entrées, ou pour désapprendre des usages (champ émergeant du « désapprentissage profond », par opposition à l'apprentissage profond – Bourtole *et al.*, 2019) tout en gardant la connaissance des usages, le règlement général sur la protection des données (RGPD) (CNIL, 2018) et les évaluations des facteurs relatifs à la vie privée (« *privacy impact assessment* ») sont des cadres et des outils à maîtriser.

Les développeurs de solutions d'IA et les entités IA doivent améliorer les algorithmes en matière de transparence, d'explicabilité, de sobriété en données, et mieux collecter et interpréter les données, ainsi que développer les modèles selon les objectifs et les données réelles et non juste les données idéalisées. Autant ces chantiers en IA sont en plein développement pour les spécialistes de l'IA, autant il est bien plus intéressant de les travailler directement avec les parties prenantes des domaines d'affaires où l'IA est appliquée, dans ce cas-ci, les travailleurs du domaine culturel. Le transfert de connaissances et de compétences entre organismes culturels est tout aussi important ; par exemple, les mesures du Plan culturel numérique du Québec (2022) et les critères du Fonds Stratégie numérique du Conseil des arts du Canada (2022) encouragent ces pratiques. D'autres organismes tiers tels que l'organisme sans but lucratif (OBNL) Synapse C, l'incubateur-accelérateur Zú ou le groupe Entertain-AI se sont créés au Canada afin de jouer le rôle d'interface entre la technologie, les données et les industries culturelles et créatives.

6. ÉVOLUTION ET ÉCONOMIE DE LA DONNÉE ET DE L'IA

Développer des solutions d'IA n'est pas à la portée de tous et nécessite la collecte de données, le développement et l'entraînement des modèles, leur évaluation et amélioration, de manière itérative et continue. De nouvelles données alimentent les algorithmes, qui sont déployés dans un nouveau contexte. Parce que les solutions d'IA influent sur la consommation, ils influent sur la distribution des revenus : ils sont prescriptifs, comme en témoigne l'influence des réseaux sociaux par rapport aux médias traditionnels.

Chacune de ces tâches nécessite du temps, de l'expertise et des ressources. Les parties prenantes (partenaires technologiques,

établissements de recherche, organismes culturels et pouvoirs publics) doivent trouver des configurations pour créer des espaces, des moments, des cadres de collaboration en matière de données, et développer des leviers financiers pour pérenniser le développement de solutions d'IA viables pour tous au niveau individuel et cultiver un écosystème viable d'organismes culturels et entités IA, et de relations entre eux sur le plan « sociétal ».

Les données et les algorithmes génèrent de la valeur et sont le fruit de travaux de multiples parties prenantes. Il devrait revenir à chacun une rémunération juste afin de pouvoir en continuer les activités de manière soutenable et viable pour tous.

CONCLUSION

Développer et déployer des solutions de l'IA nécessite de faire parler les données ensemble, ce qui constitue également un devoir et une occasion de faire parler les acteurs ensemble. Ce dialogue collectif est une condition nécessaire à l'opérationnalisation pertinente et juste de l'IA en culture, tout en impliquant en particulier les acteurs culturels dans leurs étapes de transformation organisationnelle.

Les partenariats de données sont des pratiques à la fois officielles et officieuses, dont peu de cas réels sont rapportés et étudiés par leur nature de pratique émergente ou par confidentialité en arts et en culture. Il est pertinent d'étudier le déploiement de ces modèles dans la réalité des contextes opérationnels et comment ils pourraient amener un développement et un déploiement de l'IA plus responsable, pertinent et inclusif.

RÉFÉRENCES

Instrument juridique

UNESCO, *Recommandation sur l'éthique de l'intelligence artificielle*, SHS/BIO/PI/2021/1, 23 novembre 2021.

Doctrine

Altai, « The assessment list for trustworthy artificial intelligence », 2022, <<https://altai.insight-centre.org/>>.

Anderson, Chris, « The long tail », *The Wired*, 10 janvier 2004, <<https://www.wired.com/2004/10/tail/>>.

- Born, Georgina, Jeremy Morris, Fernando Diaz et Ashton Anderson, « Artificial intelligence, music recommendation and the curation of culture », Schwartz Reisman Institute for technology and Society, University Of Toronto, 2021, <https://srinstitute.utoronto.ca/s/Born-Morris-et-al-AI_Music_Recommendation_Culture.pdf>.
- Bourtole, Lucas, Varun Chandrasekaran, Christopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie et Nicolas Papernot, « Machine unlearning », 2019, <<https://arxiv.org/abs/1912.03817>>.
- CAPACOA, « Un avenir numérique lié », 2022, <<https://linkeddigitalfuture.ca/fr/accueil/>>.
- Caramiaux, Baptiste, « The use of artificial intelligence in the cultural and creative sectors », mai 2020, <<https://research4committees.blog/2020/09/07/the-use-of-artificial-intelligence-in-the-cultural-and-creative-sectors/>>.
- Commission nationale de l'informatique et des libertés (CNIL), « Le règlement général sur la protection des données – RGPD », 23 mai 2018, <<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>>.
- Conseil des arts du Canada, « Fonds stratégie numérique », 18 août 2022 <<https://conseildesarts.ca/financement/fonds-strategiques/fonds-strategie-numerique>>.
- Culture Creates, 18 août 2022, <<https://culturecreates.com/fr/>>.
- Digital Curation Center, « The Role of Data in AI », novembre 2020, <<https://gpai.ai/projects/data-governance/role-of-data-in-ai.pdf>>.
- Estermann, Beat, « Creating a linked open data ecosystem for the performing arts », (*LODEPA Arti dello Spettacolo / Performing Arts* (Year 6), 2020, p. 31-49.
- Gagnon-Turcotte, Sarah, Miranda Sculthorp et Steven Coutts, « La gouvernance des partenariats de données numériques », *Nord Ouvert*, 2020.
- Ghosh, Paramita, « Challenges of data quality in the AI ecosystem », 2019, <<https://www.dataversity.net/challenges-of-data-quality-in-the-ai-ecosystem/>>.
- Henke, Nicolaus, Jordan Levine et Paul McNerney, « You don't have to be a data scientist to fill this must-have analytics role », *Harvard Business Review*, 2018, <<https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/analytics-translator>>.
- Kulesz, Octavio et Thierry Dutoit, « L'intelligence artificielle dans l'art et les industries culturelles et créatives. Panorama des technologies, expertises et bonnes pratiques dans l'espace francophone », 2020, <https://www.francophonie.org/sites/default/files/2020-05/OIF_Guide-IA-art_VF.pdf>.
- Latour, Bruno, *Pandora's Hope : Essays on the Reality of Science Studies*, Cambridge, Harvard University Press, 1999.
- MétaMusique, « Les meilleures pratiques pour activer la découvrabilité de votre musique », 2022, <<https://metamusique.ca/a-propos>>.
- Ministère de la Culture et des Communications du Québec, *Plan culturel numérique du Québec*, 2022, <<https://www.quebec.ca/gouvernement/ministere/culture-communications/publications/plan-culturel-numerique>>.
- Redman, Thomas C., « If your data is bad, your machine learning tools are useless », 2018, <<https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless>>.
- Ribeiro, Marco Tulio, Singh Sameer et Carlos Guestrin, « Why should I trust you? : Explaining the predictions of any classifier », *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016.

- Rogati, Monica, « The AI hierarchy of needs », *Hackernoon*, 12 juin 2017, <<https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007>>.
- SHAP, « SHAP documentation », 2022, <<https://shap.readthedocs.io/en/latest/index.html>>.
- Suresh, Harini et John Gutttag, *A Framework for Understanding Unintended Consequences of Machine Learning*, 2019, <<https://dl.acm.org/doi/fullHtml/10.1145/3465416.3483305>>.
- Tamis, « Projet Tamis. À propos », 18 août 2022, <<https://tamis.ca/a-propos/>>.
- Wilde, Damien, « Almost 90% of all uploaded YouTube videos will never reach 1,000 view », *9to5Google*, 2020, <<https://9to5google.com/2020/08/10/almost-90-of-all-uploaded-youtube-videos-will-never-reach-1000-views/>>.

PARTIE III

L'INTELLIGENCE ARTIFICIELLE ET LA SAUVEGARDE DU PATRIMOINE CULTUREL

INTELLIGENCE ARTIFICIELLE ET PRÉSERVATION DES LANGUES ET PRATIQUES CULTURELLES SÉNÉGALAISES S'ADAPTER OU DISPARAÎTRE DANS LA MONDIALISATION ?

Serigne Momar Sarr¹ et Alassane Koringho Sagna²

The rapid and widespread implementation of mobile communication networks across the region will help African countries capitalise on this late-mover advantage to adopt the most effective forms of AI in a contextually relevant manner.

Isaac Rutenberg³

Si l'Afrique est passée à côté des trois premières révolutions dites industrielles (deux industrielles et une numérique en vérité), elle ne semble guère plus entrer de bon pied dans ce qu'il est convenu d'appeler la quatrième révolution industrielle marquée par l'avènement de l'intelligence artificielle (IA). Cette situation est, toutefois, due à la fameuse « situation coloniale » (Balandier, 1951) qui a participé à la déshistoricisation des sociétés africaines. Ainsi, l'entrée symbolique de l'humanité dans l'ère de la quatrième

-
1. Enseignant-chercheur au Département de sociologie à l'Université Cheikh Anta Diop de Dakar (UCAD).
 2. Doctorant en socio-anthropologie du développement et enseignant-vacataire à l'Université Gaston Berger de Saint-Louis.
 3. Government Artificial Intelligence Readiness Index, 2019.

révolution industrielle suscite de très nombreuses interrogations, notamment au sein du continent africain. Une fois de plus, nous estimons que les cultures locales en Afrique sont interpellées dans leur double ambition de préservation et promotion, notamment en ce qui concerne les langues. L'étude de cas au Sénégal offre une analyse particulièrement édifiante sur l'avenir des cultures africaines tant le problème semble demeurer entier malgré les initiatives retentissantes comme le projet de réécriture de l'histoire générale de l'Afrique sous la conduite d'Amadou Mokhtar Mbow alors secrétaire général de l'Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO).

L'Afrique a pendant longtemps porté et porte aujourd'hui encore le poids des préjugés quant à son processus de développement en lien avec le caractère « froid » des sociétés qui la peuplent (Lévi-Strauss, 1961⁴). Suivant cette vision, elle a raté ainsi les principaux bouleversements dans le domaine de l'industrie et dans celui du numérique. Non seulement nous décelons l'« injustice épistémique » (Bhargava, 2013) que l'Afrique a toujours subie, mais nous rappelons aussi les analyses d'Appadurai (2001) sur « la production globale de la localité » au sein de l'État-nation. Cette localité concerne en fait une structure de sentiment, une propriété de la vie sociale et une idéologie d'une communauté identifiée. Les usages du numérique sont certes irréversibles, mais son adoption est soumise à des défis technologiques et des enjeux de gouvernance que la plupart des pays africains peinent à satisfaire à cause notamment des fractures et des commotions subies ainsi que des maux et des insuffisances trainées en matière de gouvernance de tous les secteurs de la vie publique.

Depuis le début de ce millénaire, des avancées considérables ont été notées dans le domaine de l'IA. Le concept de « l'homme augmenté » est là pour en témoigner. Le phénomène s'amplifie même avec les audacieuses entreprises d'Elon Musk malgré les préventions éthiques et le risque industriel. Ce champ, entendu au sens bourdieusien, dessine les contours d'un monde nouveau au sein duquel l'homme n'est pas seul détenteur de l'intelligence, mais partage celle-ci avec des machines d'un niveau de sophistication alors jamais atteint. Seulement, les scénarios ne sont pas qu'optimistes. De même qu'avec le numérique, l'IA aura de nombreuses incidences sur la marche de nos sociétés, africaines notamment. C'est tout naturellement que l'on assiste à « l'émergence d'un régime de vérité », d'une

4. Dans *Entretiens avec Claude Lévi-Strauss*, Georges Charbonnier (1961).

« *alètheia* algorithmique » (Sadin, 2021) qui préfigure un abandon de la conscience et tout bonnement la « disparition » du réel.

Les incidences de l'usage démocratisé de l'IA en Afrique auront indubitablement des répercussions, négatives et positives, sur le plan culturel au sein de nombreuses sociétés africaines. C'est tout un chapelet de croyances, d'us, de coutumes, de créations culturelles et de représentations sociales qui se verra sérieusement agiter. De nombreuses activités socioéconomiques, à l'instar de la pêche traditionnelle et de l'agriculture (les pratiques culturelles notamment), commencent déjà à subir les effets néfastes découlant des technologies liées à l'IA. Sous un registre autre, le développement de l'IA au sein des sociétés africaines va contribuer, dans une certaine mesure, au développement de « cultures intelligentes » favorisant, par exemple, la promotion des langues africaines par les supports ou les produits de l'IA.

Au Sénégal, le processus de développement numérique enclenché depuis maintenant plusieurs décennies a permis au pays de tracer une nouvelle voie vers les technologies. Bien qu'étant à un stade de balbutiement, l'IA suscite de plus en plus d'intérêt, notamment chez les jeunes chercheuses et chercheurs sénégalais.e.s qui ont d'ailleurs commencé à mettre en place des solutions intelligentes à forte influence sociale et fortement ancrées dans les réalités socioculturelles de leur pays. La promotion des langues locales est ici érigée en priorité pour faciliter l'acceptabilité sociale des applications développées. Dans le milieu de la recherche, les questions liées à l'IA sont généralement traitées sous la loupe des sciences dites dures en passant sous silence les implications sociologiques, psychologiques, anthropologiques, cognitives... de l'utilisation des technologies de l'IA dans nos sociétés. Par conséquent, le champ de recherche sur l'IA reste un vaste terrain à labourer par les sciences humaines et sociales en s'inscrivant dans une approche transdisciplinaire. Notre présent article s'attache, entre autres, à combler cette lacune. L'hypothèse de base que nous émettons est que l'IA, utilisée de manière conséquente, permettra une meilleure diffusion des savoirs et des savoir-faire grâce aux technologies et applications basées sur les langues nationales (donc une promotion de celles-ci), mais elle aura des effets négatifs sur certaines activités socioéconomiques considérées comme des pratiques culturelles chez de nombreuses ethnies sénégalaises.

Nous ferons d'abord, dans ce papier, un petit état des lieux du niveau d'évolution de l'IA au Sénégal, puis nous aborderons les représentations

sociales sur l'IA au sein de certaines ethnies sénégalaises avant de terminer par évoquer les potentiels effets positifs et négatifs des technologies de l'IA sur les langues et les pratiques culturelles de la société sénégalaise.

1. LE SÉNÉGAL À L'ÈRE DE L'IA

1.1 Développer les technologies de l'information et de la communication (TIC) pour baliser le chemin aux technologies de l'IA

Les TIC connaissent un développement fulgurant au Sénégal. Les statistiques annuelles sur l'utilisation d'Internet montrent que la société sénégalaise devient de plus en plus « connectée » au reste du monde. Cela est principalement lié à des prouesses réalisées dans le développement infrastructurel dans les domaines de la technique et du numérique. L'État sénégalais s'est engagé, depuis le dernier quart du siècle précédent, à promouvoir les TIC par l'éclosion d'un environnement numérique favorisant, entre autres, l'adoption de technologies innovantes et la dématérialisation au sein de l'administration publique.

Selon l'Observatoire sur les systèmes d'information, les réseaux et les inforoutes au Sénégal (OSIRIS), le Sénégal compterait 15 154 400 abonnés et abonnées aux services Internet en 2020. Ce chiffre représente la quasi-totalité de la population sénégalaise estimée à 17 223 497 par l'Agence nationale de la statistique et de la démographie (ANSD)⁵. Ce résultat contraste avec le fort taux d'analphabétisme de la population sénégalaise. Mais cela ne constitue pas véritablement un frein à l'usage d'Internet et de ses applications puisque les analphabètes sénégalais et sénégalaises ont su développer des « compétences numériques » (*digital skills*) leur permettant d'utiliser les outils TIC (Jimbira et Cissé, 2018).

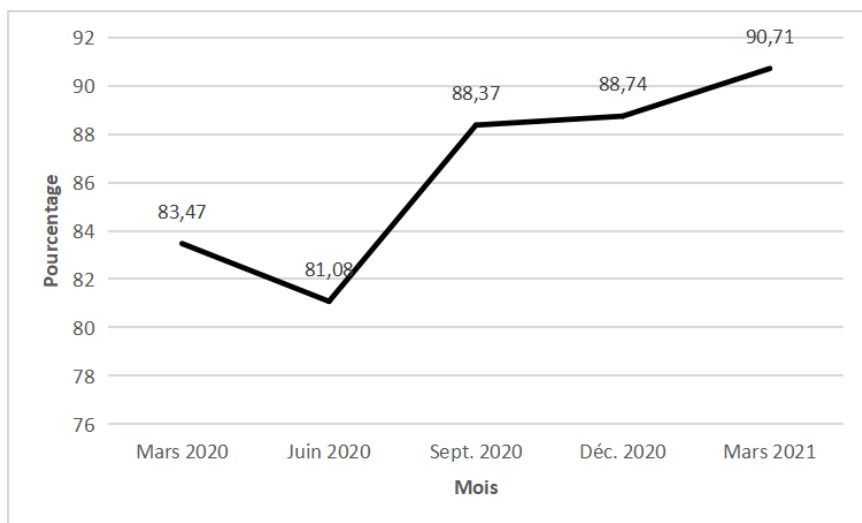
L'autre paramètre essentiel pour évaluer le niveau d'évolution d'un pays en matière de TIC est le taux de pénétration d'Internet. Entre 2020 et 2021, ce taux a connu une importante évolution au Sénégal, passant ainsi de 83,47 % en mars 2020 à 90,71 % en mars 2021, soit une hausse de plus de 7 % en une année comme l'atteste la figure 12.1⁶. Naturellement, l'accès aux services Internet est assez faible dans les milieux ruraux sénégalais en raison d'une faible couverture réseau (couverture uniquement

5. Projection de la population en 2021.

6. Données fournies par l'Observatoire Internet de l'Autorité de régulation des télécommunications et postes (ARTP), mars 2021.

en réseau 2G/3G) alors que dans les milieux urbains la couverture reste plutôt satisfaisante avec le réseau 4G.

FIGURE 12.1
Évolution trimestrielle du taux de pénétration d'Internet au Sénégal



Mars 2020 à mars 2021

Source: Autorité de régulation des télécommunications et postes (ARTP, 2021).

Les chiffres évoqués nous permettent d'appréhender le niveau de développement numérique au Sénégal où la démocratisation d'Internet, nécessaire pour assurer une révolution numérique, est devenue une réalité. Au regard de ce qui précède, d'aucuns seraient tentés de se poser, à raison d'ailleurs, la question suivante : quel est donc le lien entre IA et TIC ? Nous pouvons ici considérer l'IA comme une sophistication complexe des TIC. Son développement nécessite préalablement une bonne maîtrise des outils TIC qui se trouve être le principal préalable pour ouvrir une porte nouvelle menant à ce nouveau « monde intelligent » en construction.

Si nous ne pouvons affirmer, avec certitude, que le Sénégal a réussi sa révolution numérique, nous pouvons, au moins, retenir que des avancées significatives sont à mettre à son compte. Les statistiques sur le taux de pénétration d'Internet, le nombre d'abonné.e.s aux services Internet, l'utilisation des outils TIC dans les activités socioprofessionnelles, l'enseignement des TIC à l'école et dans les universités, etc., sont plutôt satisfaisantes. Fort de ces avancées, le pays a commencé à s'intéresser de plus

en plus à l'IA dans ses dimensions à la fois théoriques (création de formations dédiées à l'IA) et pratiques (développement de solutions à fortes répercussions sociales basées sur l'IA). Enfin, la prolifération des entreprises en démarrage (*startups*) est un indicateur sur l'usage croissant et exponentiel qui laisse présager un développement fulgurant au cours des années à venir.

1.2 Évolutions en cours dans le domaine de l'IA au Sénégal

D'importantes évolutions sont en cours dans le domaine de l'IA au Sénégal. Ces évolutions se mesurent à l'aune des multiples solutions mises en place par de jeunes Sénégalais et Sénégalaises pour influencer la marche de leur société. De l'assistant vocal en langue locale (*Wolof*)⁷ à l'interface numérique dédiée aux personnes en situation de handicap, les solutions créées intègrent une dimension culturelle avec un accent mis sur la promotion des langues locales. Tenir compte des langues locales permet une plus large et meilleure utilisation de ces solutions proposées par de jeunes chercheurs et chercheuses.

Dans le domaine de l'enseignement supérieur, des programmes de formations en IA sont en cours de mise en œuvre dans les établissements privés et publics. Ces établissements, à l'instar de l'Université virtuelle du Sénégal (UVS), forment des étudiants et étudiantes capables de créer des applications basées sur l'IA avec un accent particulier mis sur l'effet social de ces applications. L'application Wax App (« Wax » qui signifie « parler » en wolof, principale langue parlée au Sénégal) par exemple, disponible sur les plateformes de téléchargement, permet aux personnes en situation de handicap, notamment les sourds-muets, de communiquer, une fois l'application installée sur un téléphone intelligent (*smartphone*), avec le reste de la société grâce à une voix de synthèse en langue locale. Cela réduit considérablement l'exclusion sociale dont sont souvent victimes ces personnes.

Les chercheuses et chercheurs formés tentent d'établir un lien entre les solutions à mettre en place et les réalités socioculturelles sénégalaises. Les langues locales comme le wolof, le pular, ou encore le sérère sont ainsi mises à contribution pour accroître l'acceptabilité sociale et faciliter l'utilisation des applications développées. Dans le domaine de l'agriculture, par exemple, des élèves-ingénieurs de l'École Supérieure Polytechnique (ESP)

7. Il s'agit d'une solution développée par une étudiante de l'Université Gaston Berger (UGB) de Saint-Louis. L'assistant vocal permet de communiquer avec une maison dite connectée.

de Dakar ont mis en place une solution innovante basée sur l'IA nommée Tol bi (le champ en wolof). Il s'agit d'une plateforme de traitement d'images, par le biais de l'IA, pour la collecte, le traitement et la prise de décision. À l'aide d'un simple téléphone intelligent, le paysan a accès, à distance, aux informations sur son périmètre agricole (maladies des plantes, pertes en eau, type d'engrais à utiliser, etc.). Le paysan est guidé par une voix qui s'exprime en langue locale (le wolof) pour une appropriation et une meilleure utilisation de l'application Tol bi. Selon ses créateurs, Tol bi a contribué à optimiser jusqu'à 30 % des rendements agricoles grâce aux informations transmises aux paysans.

Nonobstant certaines prouesses, il reste beaucoup à faire principalement dans le travail collaboratif qui peine à voir le jour. En fait, les solutions développées sont souvent le fruit d'efforts individuels et ne bénéficient généralement pas de l'appui du gouvernement sénégalais. En vérité, très peu de financements sont destinés à la recherche, tous domaines confondus. Pour l'IA, aucun fonds n'a encore été mobilisé par l'État du Sénégal pour financer la recherche dans ce domaine. Cela constitue un des principaux freins au développement de l'IA. Malgré donc la détermination remarquable des jeunes ingénieurs sénégalais.e.s, il manque une vision stratégique, en plus de l'absence de financement, qui doit être pensée et conçue par l'État en collaboration avec l'ensemble des acteurs et actrices du domaine de l'IA. On comprend, *in fine*, que le Sénégal est un pays qui n'a pas totalement les moyens de son développement.

Le peu d'intérêt manifesté par les pouvoirs publics sénégalais à l'égard de l'IA n'empêche pas les jeunes ingénieurs et ingénieures et élèves ingénieurs et ingénieures à développer leurs connaissances et compétences dans le domaine de l'IA. À défaut d'une mise en pratique de ces connaissances et compétences au niveau local, ces jeunes s'expatrient pour faire valoir leur savoir-faire ailleurs. Moustapha Cissé, par exemple, après des études au Sénégal puis en France, a su gagner la confiance de Google au point d'être nommé à la tête du premier centre africain de recherche en intelligence artificielle⁸. Pour Elias W. Ba, président de l'Association sénégalaise de l'intelligence artificielle (Galsene IA)⁹, « ce qui manque actuellement, c'est une réussite collective dans le domaine de l'intelligence

8. Moustapha Cissé a récemment quitté ce poste pour se lancer dans une aventure individuelle « au service du continent africain », dit-il.

9. Galsene IA est une association regroupant plus de 1 000 membres passionnés d'IA. Elle est un espace de dialogue et de formation qui vise à promouvoir l'IA au Sénégal.

artificielle». Cette réussite collective est, en partie, liée à l'engagement de l'État dans la création et la promotion de formations et de recherches sur l'IA. Comme souligné précédemment, l'IA souffre de l'absence d'une vision stratégique de l'État et d'une absence totale de financement de la recherche dans ce domaine. Notons que ce n'est que récemment que le Sénégal a pris de forts engagements dans le domaine du financement de la recherche scientifique et technique avec, certes, des moyens peu importants. Quant à l'IA, elle reste confinée dans une tour d'ivoire, laissée à elle-même ou plutôt à ceux et celles qui ambitionnent de la développer.

Les évolutions en cours dans le domaine de l'IA au Sénégal ouvrent un boulevard d'espoirs quant au développement rapide de cette technologie nouvelle. Dans un avenir peut-être encore lointain, les solutions basées sur l'intelligence artificielle seront au cœur de la vie sociale des Sénégalais et Sénégalaises. Cela invite à penser le rapport entre culture, représentations sociales et IA dans un contexte social et culturel différent de celui de la naissance de cette quatrième révolution industrielle.

2. RAPPORT ENTRE CULTURE, REPRÉSENTATIONS SOCIALES ET IA CHEZ CERTAINES ETHNIES SÉNÉGALAISES

Discuter de la culture revient avant tout à tracer les contours sémantiques de ce concept aux multiples définitions qui ne manquent pas de produire de la confusion très souvent. Nous nous suffirons, pour ce sujet précis, de la définition de la culture donnée par Taylor (1871) qui est, selon lui, «ce tout complexe comprenant à la fois les sciences, les croyances, les arts, la morale, les lois, les coutumes et les autres facultés et habitudes acquises par l'homme dans l'état social». Dans le cas sénégalais qui nous préoccupe ici, nous utiliserons cette notion au pluriel (les cultures) pour relever les particularités de chaque ethnie sénégalaise en matière d'identité et de créations culturelles.

La croyance en des objets habités par une âme ou une intelligence n'est pas un fait nouveau pour une bonne partie de la société sénégalaise, notamment chez les *Joola* et les *Sérère*. Les peuples de ces ethnies du Sénégal ont depuis longtemps cru et continuent de croire aujourd'hui au caractère sacro-saint et aux pouvoirs surhumains de certains membres de leur communauté qui sont généralement désignés sous le nom de «sages». Ceux-ci auraient, par des pouvoirs mystiques, surclassé l'intelligence humaine et ont su se démarquer de leurs semblables en société par leur capacité à créer un lien entre le monde des humains et celui des ancêtres,

des dieux et même des non-humains. Les « prouesses mystiques » réalisées par ces « supra-hommes », n'entrent dans aucune considération cartésienne et ressemblent plus à des mythes ou légendes qu'à du factuel pouvant être soumises à un examen rationnel. Pour autant, si l'existence des « prouesses mystiques » évoquées ne peut être réellement démontrée auprès d'un esprit cartésien, le contraire ne peut non plus l'être de façon scientifique.

Les « prouesses mystiques », dont il est ici question, passent par l'établissement de liens entre le monde des humains et celui des dieux au moyen d'une sorte de « connexion » faite par les « sages » qui ont fini de transcender les modes de raisonnement et l'intelligence de l'homme en tant qu'agent social. Dans cette sorte de « connexion », l'élément moteur servant de pont entre deux mondes différents est un arbre, un animal ou un quelconque autre objet de la vie sociale. Chez certaines populations de l'ethnie *Joola* dans le sud du Sénégal, les animaux, ou plutôt certains d'entre eux, sont « habités » par une âme qui fait qu'ils sont, dans une certaine mesure, dotés d'intelligence. Cette idée est très répandue dans d'autres ethnies sénégalaises qui, malgré la prégnance des religions monothéistes (islam et christianisme), pratiquent toujours l'animisme. Selon Philippe Descola¹⁰, l'animisme « attribue à tous les êtres humains et non humains le même genre d'intériorité, de subjectivité, d'intentionnalité. Il place la différence du côté des propriétés et manifestations physiques : apparence, formes du corps, manière d'agir, comportements ». Les représentations présentes dans le champ de l'animisme ne semblent pas très éloignées de celles du champ de l'IA appréhendée, plus spécifiquement, sous l'angle de la relation homme-machine avec naturellement quelques lignes de démarcation.

Pour revenir à la relation entre « sages » et divinités ou ancêtres d'outre-tombe chez les peuples *Joola*, retenons qu'ici l'établissement d'une communication passe par un élément moteur (un animal « intelligent ») pour atteindre les récepteurs (les dieux) et leur délivrer le message envoyé par le commun des mortels. Chez les *Lebbu* de la presqu'île du Cap-Vert, il existe aussi cette croyance en un pouvoir intermédiaire situé entre l'espace temporel, donc terrestre, et l'espace spirituel (en rapport avec les divinités ou les ancêtres). Ici également, les messages du commun des mortels sont délivrés par des « sages » eux aussi dépositaires d'impressionnants pouvoirs mystiques.

10. Entretien avec Nicolas Journet, *Sciences Humaines*, Grands Dossiers n° 5, décembre 2006, janvier-février 2007.

Partant de ce qui précède et si nous nous référons aux représentations socio-symboliques comprises dans l'animisme, nous pouvons retenir que l'idée d'une intelligence en dehors de l'humain (présente chez un objet par exemple) n'est pas si étrangère chez de nombreuses ethnies du Sénégal dont une partie pratiquent encore les cultes animistes. Dans l'animisme, on va même au-delà de la présence de la faculté « raison » ou « intelligence » chez des objets ou des animaux. Il est, la plupart du temps, question de sensibilité, d'intention, d'empathie, d'interaction, de compréhension, etc., entre les animaux ou les objets habités par une « âme » et les « supra-humains » qui établissent une relation communicationnelle entre le monde terrestre et l'autre monde. Chez les robots intelligents, les capacités à sentir ou à comprendre le monde extérieur restent à développer et à affiner mieux encore pour espérer des résultats satisfaisants à long terme.

Dans le fond, les représentations que nous pouvons retrouver chez les animistes et certains principes de fonctionnement de l'IA ont plusieurs points en commun. Seulement, toute la différence réside ici dans le vocable « artificiel ». Chez les ethnies citées plus haut, l'intelligence est plus d'ordre naturel, elle n'émane point d'une volonté humaine et elle prend naturellement possession d'un objet ou d'un animal. S'agissant de l'IA que nous connaissons aujourd'hui, il est bien question d'une intelligence « artificielle », c'est-à-dire fabriquée par une autre intelligence : l'intelligence humaine. Les mains des dieux ou des ancêtres derrière l'intelligence « naturelle » connue chez la plupart des ethnies sénégalaises et la main de l'homme derrière l'IA, voilà qui résume bien cette dichotomie. Donc, « la guerre des intelligences » (Alexandre, 2017) aura bien lieu.

A priori, les technologies de l'IA ne sauraient véritablement étonner certains peuples d'ethnies dont les pratiques religieuses sont orientées vers l'animisme. Néanmoins, le caractère réel et plus factuel de ces nouvelles technologies (une voiture intelligente par exemple) peut amener ces peuples à réinterroger et à repenser leurs pratiques culturelles et leur logique d'appréhension du réel. Le déploiement des technologies de l'IA dans la société sénégalaise (prise comme une agrégation complexe d'ethnies et de cultures) ne peut donc se faire sans conséquences. Les principales répercussions sur lesquelles nous mettons l'accent ici sont en lien avec les langues locales et certaines pratiques culturelles ancestrales.

3. LES DÉFIS DES LANGUES ET DES PRATIQUES CULTURELLES SÉNÉGALAISES FACE À L'IA GÉNÉRALISÉE

3.1 IA : menace ou occasion pour la préservation des langues locales dans un nouveau monde intelligent ?

Le contexte de mondialisation, qui implique une rencontre des cultures, était un prétexte pris par de nombreux États pour travailler à préserver leurs langues et créations culturelles et échapper ainsi au diktat d'une culture-monde. Avec le passage d'un village planétaire, dominé par le numérique, à un monde intelligent dominé par l'IA, la question de la préservation des langues locales s'imposera plus encore dans les débats publics, particulièrement dans des contrées présentant de larges diversités linguistiques. Au Sénégal précisément, les progrès dans le domaine de l'IA vont, sans doute aucun, avoir des répercussions sur les dynamiques de préservation et de vulgarisation de savoirs et de savoir-faire pour les langues du pays.

Les répercussions de l'IA sur les dynamiques de préservation des langues locales seront en grande partie positives si toutefois il est fait bon usage des outils et applications développés grâce à cette technologie nouvelle. Les acteurs sénégalais travaillant sur l'IA avec lesquels nous nous sommes entretenus mettent en avant l'importance de la préservation des langues locales pour valoriser et pérenniser les identités culturelles dans un contexte de mondialisation grandissante. Précisons, à cet effet, que les langues sont un puissant instrument qui véhicule l'identité culturelle et les créations culturelles d'un peuple. Mieux encore, elles représentent le noyau dur d'une culture avec comme éléments périphériques toutes les autres dimensions de la culture. Par conséquent, retenons que c'est par la langue qu'un peuple se forme et se transforme (par le métissage culturel), certes à partir d'aires culturelles. Face au défi de la préservation des langues locales dans un monde intelligent en construction, une bonne réappropriation de l'IA devrait permettre au Sénégal de préserver et même de vulgariser plus efficacement ses langues. Soulignons le fait que l'appropriation des technologies de l'IA dépend largement de sa tropicalisation, c'est-à-dire son adaptation aux réalités socioculturelles africaines (ici sénégalaises).

Par ailleurs, le développement des technologies de l'IA, arc-bouté sur les langues locales, pourrait aider à résoudre l'épineuse question de l'analphabétisme dans un pays où la langue d'enseignement est étrangère (le

français), ce qui constitue un handicap dans l'accès aux savoirs¹¹. Elles offriront l'avantage de faciliter l'accès aux savoirs littéraires, scientifiques, techniques, etc., à travers les langues parlées par les diverses ethnies qui peuplent le Sénégal. Dans le contexte sénégalais, « l'intelligence artificielle peut aider à bâtir des modèles pédagogiques de formations pour accélérer l'accès des femmes, des hommes, des jeunes filles et des garçons à l'écriture, à la lecture et à l'utilisation des langues nationales » précise le professeur Mary Teuw Niane¹².

L'accélération de l'accès aux diverses formes de savoirs et savoir-faire passera, entre autres, à l'ère de l'IA, par l'entremise de « robots-enseignants ». Ces robots-enseignants sont généralement équipés d'un écran-tablette et dotés d'un système de reconnaissance faciale et de capteurs. Ces dispositifs techniques leur permettent d'interagir avec des élèves et des enseignants et enseignantes. Au regard des expériences déjà réalisées et de celles en cours, nous pouvons affirmer que la transmission de connaissances par un robot humanoïde pourrait se révéler plus efficiente qu'une transmission portée par un être proprement humain. En effet, le robot en question est capable de déceler, au moyen d'un test de quelques minutes, les capacités et aptitudes des apprenants et apprenantes et de choisir le meilleur programme d'enseignement qui convient à chacun.

En 2019, à Nantes, un « robot-enseignant » a donné plusieurs heures de cours à des élèves ingénieurs de l'École centrale. Les résultats sont jugés plutôt satisfaisants comparés à ceux d'une expérience humaine d'enseignement. À l'avenir, déployer des « robots-enseignants » programmés sur la base des langues locales dans le contexte sénégalais peut aider à réduire considérablement le taux d'analphabétisme et faciliter la transmission de connaissances dans les langues du pays. Toutefois, comme ailleurs dans le monde, mais ici plus encore, cela ne se révèle pas être d'emblée une évidence puisque se posera la question de l'acceptabilité sociale de cette technologie qui s'imposera dans le nouveau monde intelligent. Déjà en Europe, des organismes ou associations, à l'instar du Comité syndical européen de l'éducation (CSEE), mènent des campagnes contre l'expérimentation, à tout

11. Cf. les travaux de Cheikh Anta Diop sur les langues et cultures africaines dans son ouvrage *Nations nègres et cultures* (1979).

12. Mary Teuw Niane est professeur titulaire de mathématiques de classe exceptionnelle. Ancien recteur de l'Université Gaston Berger de Saint-Louis, il est également ancien ministre de l'Enseignement supérieur du Sénégal et a été la « cheville ouvrière » des réformes de l'enseignement supérieur au Sénégal et est à l'origine de la création de l'Université virtuelle du Sénégal. Le professeur Niane est très engagé dans les questions liées à l'IA.

va, de l'éducation numérique, à travers des robots, dans les écoles. De l'avis du CSEE, « les robots ne remplaceront jamais les enseignant(e)s¹³ ». Voilà qui semble être déjà un présage de mauvais augure pour l'introduction de robots-enseignants dans le domaine de l'éducation au Sénégal.

S'adapter au nouveau monde intelligent en construction notamment sur le plan linguistique sera un défi à relever par la société sénégalaise. Si les technologies de l'IA sont promptes à appuyer les dynamiques de préservation et valorisation des langues locales, elles peuvent, en revanche, constituer des menaces pour celles-ci. En vérité, si l'on part du fait que l'IA est elle-même une culture (qui a les attributs d'une « culture monde »), elle peut participer à l'élargissement du fossé déjà existant entre différentes cultures du monde avec un centre dominant et plusieurs périphéries dominées. Sans une réelle appropriation et une adaptation des technologies et des solutions basées sur l'IA à ces réalités socioculturelles, la société sénégalaise, considérée ici dans sa diversité culturelle, sera soumise à la toute-puissance de la culture imposée par le nouveau monde intelligent, une culture de l'uniformité et non de la diversité.

3.2 Répercussions de l'IA sur les pratiques culturelles sénégalaises

Le développement de l'IA au Sénégal permettra d'élargir le pont menant à la culture-monde en facilitant ainsi les diffusions de pratiques et de créations culturelles. Le fameux « rendez-vous du donner et du recevoir » théorisé par l'ancien président sénégalais Léopold Sédar Senghor s'en verra renforcé davantage. Pour le professeur Maïssa Mbaye¹⁴ : « Nous sommes dans une phase transitoire où nous subissons la culture et je pense qu'avec l'intelligence artificielle appliquée à nos problèmes et nos réalités au niveau local cela peut aider à montrer que le Sénégal a quelque chose à apprendre au monde en termes de cultures ». Dans cette logique, l'ouverture des cultures locales (sénégalaises) à la culture-monde passera par un mouvement de promotion des pratiques culturelles endogènes à travers des solutions typiquement fondées sur l'IA. Pour ce faire, il faudra préalablement travailler à développer les technologies de l'IA en partant des particularités et des réalités des cultures locales, sans quoi aucune

13. Cf article plus complet : Comité syndical européen de l'éducation, « Les robots ne remplaceront jamais les enseignant(e)s », 1^{er} mars 2018, <<https://www.csee-etu.org/fr/actualites/2493-teachers-will-never-be-replaced-by-robots-3>> (consulté le 20 octobre 2021).

14. Le professeur Maïssa Mbaye est directeur du Centre d'excellence africain en mathématiques et TIC (CEAMITIC) à l'Université Gaston Berger de Saint-Louis.

préservation ni promotion de pratiques culturelles ne sera possible dans le monde intelligent en construction. À ce propos, des avancées sont en cours dans le domaine de l'élevage, notamment avec le développement de solutions basées sur l'IA et tenant compte des réalités des cultures locales.

En effet, l'élevage est une pratique culturelle ancestrale chez les *Peuls* au Sénégal. Depuis plusieurs décennies, cette ethnie est aux prises avec un problème du vol de bétail et de l'intrusion du bétail dans des périmètres agricoles des populations d'autres ethnies, notamment les *Sérères*. Le dernier fait évoqué conduit très souvent à des conflits sociaux entre les éleveurs *peuls* et les agriculteurs au Sénégal ou ailleurs dans la sous-région. La solution proposée par de jeunes sénégalais et sénégalaises permet de juguler ces conflits interethniques. Grâce à l'Internet des objets, que nous considérons comme une des sections de l'IA, les éleveurs peuvent mieux contrôler leur bétail grâce à des « enclos intelligents » et des « colliers intelligents ». Au moyen d'alarmes sonores et d'alertes SMS, ces deux solutions permettent de réduire considérablement le vol de bétail ainsi que l'intrusion du bétail dans des périmètres agricoles. Elles revêtent un double intérêt : un intérêt sur le plan économique avec une réduction des pertes financières relatives au vol de bétail et un intérêt sur le plan social en permettant de mettre fin aux nombreux conflits interethniques aux conséquences souvent dramatiques. C'est donc en partant d'un diagnostic de problèmes sociaux que ces jeunes Sénégalais et Sénégalaises, regroupés autour de la plateforme Africa Smart Citizens, ont pu développer des solutions intelligentes à fortes répercussions sociales.

Les technologies liées à l'IA promeuvent, pour la plupart, un allègement des travaux réalisés par des humains et surtout une rapidité dans l'exécution de certaines tâches, ce qui revient donc à remplacer la force humaine par un outil beaucoup plus intelligent et puissant. Un robot intelligent, par exemple, est capable d'effectuer en quelques minutes une tâche qu'accomplit généralement un humain en plusieurs heures. Dans le cas de la société sénégalaise où certaines activités socioprofessionnelles ont un fort ancrage socioculturel, à l'instar de la pêche qui est une activité plus culturelle qu'économique chez les *Lebbu* de la presqu'île du Cap-Vert, le déploiement futur de robots intelligents, à l'instar du robot pêcheur chinois, mobilisés pour mener des activités halieutiques constituerait, à long terme, une menace réelle à l'activité de pêche chez ces peuples. C'est donc presque une institution sociale ancestrale (l'activité de pêche) chez les *Lebbu* qui sera menacée par l'IA. Soulignons qu'à l'heure actuelle certains bateaux de pêche étrangers utilisent déjà des technologies intelligentes,

sonars et sondeurs, pour détecter les zones à forte présence de poissons dans les eaux sénégalaises. Il y a même l'utilisation des drones pour la pêche de gros poissons ! Cette pêche intelligente renforce le risque d'épuisement des ressources halieutiques et met en péril la pratique culturelle qu'est la pêche. L'utilisation de technologies plus intelligentes et d'un niveau de sophistication jamais atteint, comme le robot pêcheur chinois qui permet de localiser et d'attirer les poissons, ne fera guère évoluer positivement cette situation.

De même que l'activité de pêche, l'agriculture est une pratique culturelle chez de nombreuses ethnies sénégalaises. Elle a certes une dimension économique, mais sa dimension sociale est ici de loin plus prépondérante. En fait, les pratiques culturelles traditionnelles chez les peuples de l'ethnie *Joola* par exemple, actifs dans la riziculture depuis plusieurs siècles, sont transmises de génération en génération. Elles constituent l'héritage culturel qui vient s'agglutiner aux nombreuses pratiques culturelles qui entourent l'activité agricole (cérémonies de prières pour demander aux dieux ou ancêtres une pluie abondante, la protection des cultures face aux invasions acridiennes, une bonne récolte, etc.). Ces pratiques tournant autour de l'activité agricole demeurent chez plusieurs ethnies sénégalaises malgré le développement, de plus en plus affirmé, de l'agriculture industrielle généralement portée par de grosses sociétés étrangères. Cette agriculture industrielle menace l'agriculture familiale qu'il ne faudrait pas voir comme un simple moyen de subsistance, mais comme un moyen de conservation de pratiques culturelles et culturelles ancestrales. De plus, elle menace le foncier local avec le phénomène d'acquisition massive de terres (*land grabbing*).

Les technologies de l'IA (robot de traite, drones agricoles, tracteurs autonomes, etc.), déployées massivement dans le domaine agricole, mèneront à une disparition de ces pratiques culturelles et culturelles avec un remplacement, peut-être inévitable dans ce domaine, de l'homme par la machine, comme ce fut le cas avec les premières révolutions industrielles. Le remplacement déjà programmé de l'homme par des machines intelligentes emportera avec lui toutes ces activités socio-symboliques autour de l'agriculture (cérémonies préhivernales de prédiction des récoltes, consultation des ancêtres pour le choix des meilleures pratiques culturelles, cérémonies de prières pour une bonne pluviométrie, par exemple). Les technologies de l'IA permettent aujourd'hui de prédire la qualité des récoltes, d'aider à choisir la meilleure pratique culturelle pour un périmètre donné et même de provoquer des pluies artificielles à travers un

réseau intelligent de générateurs au sol. Ces tâches alors confiées, dans de nombreuses ethnies du Sénégal (*Sérère* et *Joola* notamment), à des êtres «suprêmes» ou divinités d'un autre monde sont aujourd'hui aisément réalisables avec de petits gadgets intelligents.

Au-delà des pratiques culturelles, nous pouvons affirmer que les technologies de l'IA auront des conséquences négatives sur les relations sociales au sein de différentes ethnies avec un renforcement de l'individualisme qui est d'ailleurs un processus en cours avec les progrès numériques. En fait, l'utilisation presque généralisée des TIC (téléphonie mobile, Web, réseaux sociaux numériques, etc.) au sein de la société sénégalaise contribue au développement de l'individualisme dans une société qui a toujours accordé une valeur singulière et sacrée aux relations sociales et aux interactions quotidiennes. Les technologies de l'IA, avec l'idée mise en avant de création de relations «amicales» homme-machine, ne feront que renforcer cet individualisme perçu ici comme une sorte de maladie qui a fini de «ravager» l'Occident. La question du «vivre ensemble avec des êtres artificiels intelligents» (Kiyindou, 2019) va donc se poser avec acuité et sera peut-être à l'origine de réticence ou refus au sein de nombreuses sociétés quant à l'adoption des outils de l'IA.

CONCLUSION

Tout bien considéré, retenons que le développement de l'IA n'aura pas les mêmes conséquences ou incidences au sein des différentes sociétés d'Afrique ou de l'Occident. La configuration socioculturelle n'est pas la même ici et là-bas, les croyances, les us et les coutumes sont également grandement différents. Dans le cas sénégalais, qui nous a ici préoccupés, nous avons vu que les technologies développées sur la base de l'IA auront des répercussions à la fois positives et négatives sur le plan culturel. Les répercussions positives seront plus significatives, à notre sens, dans le domaine de la promotion de l'alphabétisation, qui reste un défi énorme à relever pour le Sénégal, à travers des applications en langues locales ou nationales. L'IA permettra ainsi une diffusion rapide et efficace des savoirs et savoir-faire dans les langues de chacune des ethnies peuplant le Sénégal. À l'opposé, certaines activités socioéconomiques portant l'estampe de pratiques culturelles, comme l'agriculture ou la pêche, viendront à être menacées de disparition, et le mot n'est pas fort, au fil du temps et les relations sociales se verront sérieusement bouleverser par l'essor des technologies de l'IA.

S'adapter ou disparaître face au développement des technologies de l'IA? La question ne reste pas moins actuelle malgré l'inscription de la fameuse quatrième révolution industrielle dans un univers futuriste au sein de certains pays qualifiés de sous-développés. Une telle considération s'explique par une incompréhension de tous les enjeux qui sont derrière les progrès technologiques en rapport avec l'IA. Malgré l'écart de niveau de développement avec l'Occident, des pays africains, à l'image du Sénégal, devront rapidement se mettre sur le bloc de départ de cette nouvelle révolution industrielle en cours et relever les défis de l'adaptation sur le plan culturel. Dans le monde intelligent qui se construit, les cultures qui n'arriveront pas à s'adapter aux technologies de l'IA seront sans doute vouées à disparaître ou, au meilleur des cas, à jouer un rôle parasite de réceptrices d'autres valeurs et pratiques culturelles d'ailleurs.

Aujourd'hui, les enjeux de l'IA sont si stratégiques qu'aucune nation ne peut se donner le loisir de considérer ce domaine comme non prioritaire, bien que son développement requière des ressources humaines et surtout financières assez conséquentes. Le passage en cours d'un village planétaire (avec le numérique) à un monde intelligent (avec l'IA) doit interpeller tous les États du monde et ceux d'Afrique en particulier. C'est à ceux qui se battront pour faire partie de la gouvernance de ce futur monde intelligent qu'appartient l'avenir de cette humanité, des sociétés et de leurs cultures. Pour éviter toute fâcheuse mésaventure et une perte des identités culturelles sous l'ère de l'IA, les sociétés africaines devront comprendre les enjeux et les défis de cette nouvelle révolution avant de se l'approprier. Une mauvaise maîtrise des technologies de l'IA et une absence d'anticipation sur leurs répercussions socioculturelles sont semblables à un saut dans le vide qui risque de maintenir le continent africain à la périphérie du monde, de ce nouveau monde intelligent, comme il a été le cas lors des siècles de traite et de colonialisme.

Sur le plan de la recherche en sciences humaines et sociales, il sied de contribuer à enrichir davantage la littérature mondiale sur les technologies de l'IA. En Afrique francophone, les recherches dans ce domaine restent l'apanage des sciences dites dures (mathématiques, physique, etc.). Les sciences humaines et sociales tardent encore à s'intéresser à la quatrième révolution industrielle alors que des disciplines, comme la sociologie, se sont très vite saisies des questions humaines et sociales induites par les deux premières révolutions industrielles, les techniques et le non-humain. Ne dit-on d'ailleurs pas que la sociologie est fille de la révolution industrielle britannique et de la révolution sociopolitique française du

XVIII^e siècle ? La sociologie devrait-elle, comme l'anthropologie, repenser à l'aune des réalités négro-africaines pour comprendre et se conformer aux perspectives de vie et d'avenir.

Au même titre que les pays anglophones d'Afrique qui sont assez prolifiques dans la production et la constitution de ce que nous appelons une « littérature humaine et sociale » sur l'IA, les pays d'Afrique francophone devront eux aussi faire des efforts pour faire avancer les recherches portant sur les technologies de l'IA en sciences humaines et sociales. La recherche et la production scientifique, rappelons-le, sont des éléments essentiels de la culture dite moderne. Les technologies de l'IA ne peuvent évoluer en toute ignorance des réalités socioculturelles des différentes sociétés. Sur ce point précis, seules les sciences humaines et sociales peuvent véritablement aider à appréhender les interrelations entre IA et cultures des sociétés diverses, encore qu'elles doivent être situées et ancrées d'une part et, d'autre part, vigilantes et critiques.

RÉFÉRENCES

- Alexandre, Laurent, *La guerre des intelligences : intelligence artificielle versus intelligence humaine*, Paris, JC Lattes, 2017.
- Appadurai, Arjun, *Après le colonialisme : les conséquences culturelles de la globalisation*, Paris, Payot, 2001.
- Balandier, Georges, « La situation coloniale : approche théorique », *Cahiers internationaux de sociologie*, vol. 11, 1951, p. 44-79.
- Bhargava, Rajeev, « Pour en finir avec l'injustice épistémique du colonialisme », *Socio*, n° 1, 2013, p. 41-75.
- Charbonnier, Georges, *Entretiens avec Claude Lévi-Strauss*, Paris, Plon Julliard, 1961.
- Comité syndical européen de l'éducation, « Les robots ne remplaceront jamais les enseignant(e)s », 1^{er} mars 2018, <<https://www.csee-etuce.org/fr/actualites/2493-teachers-will-never-be-replaced-by-robots-3>> (consulté le 20 octobre 2021).
- Diop, Cheikh Anta, *Nations nègres et cultures : de l'Antiquité nègre égyptienne aux problèmes culturels de l'Afrique noire d'aujourd'hui*, Paris, Présence africaine, 1979 [1954].
- Jimbira, Mohamed Sakho et Hadji Bangali Cissé, « L'usage d'Internet dans les classes populaires sénégalaises. Le cas de marabouts, marchands ambulants et femmes de ménage », *Réseaux*, vol. 2-3, n°s 208-209, 2018, p. 173-193.
- Kiyindou, Alain, *Intelligence artificielle : pratiques et enjeux pour l'Afrique*, Paris, L'Harmattan, 2019.
- Sadin, Éric, *L'intelligence artificielle ou l'enjeu du siècle : anatomie d'un antihumanisme radical*, Paris, Éditions L'Echappée, 2021.
- Taylor, Edward Burnett, *Primitive Culture*, Cambridge, Cambridge University Press, 1871.

ARCHIVES ARCHÉOLOGIQUES, INTELLIGENCE ARTIFICIELLE ET SCIENCES PARTICIPATIVES

LE PROJET « BULLIOT, BIBRACTE ET MOI »

Emmanuelle Perrin¹ et Philippe Chassignet²

L'intelligence artificielle offre de nouvelles solutions pour automatiser la transcription des manuscrits et réactualiser des archives scientifiques encore peu exploitées, difficilement accessibles et restées en grande partie inédites. Le projet « Bulliot, Bibracte et moi », lauréat de l'appel « Service numérique innovant 2019 » du ministère français de la Culture et de la Communication, constitue une expérience originale de science participative associée à l'intelligence artificielle pour la transcription d'archives archéologiques. Il vise l'édition numérique des carnets de fouille manuscrits de Jacques-Gabriel Bulliot (1817-1902). Ce négociant en vin d'Autun localisa le site de la ville celtique de Bibracte et il fut le premier à y diriger des fouilles pendant une trentaine d'années, entre 1867 et 1895. Porté par le laboratoire lyonnais Archéorient³ (Université Lumière Lyon 2 – CNRS), ce projet réunit le Centre archéologique européen de Bibracte⁴, le Musée

-
1. Docteure en histoire, a participé à la coordination scientifique du projet « Bulliot, Bibracte et moi » au sein du laboratoire Archéorient, à la Maison de l'Orient et de la Méditerranée (Lyon).
 2. Archéologue, a fait partie de l'équipe de coordination scientifique du projet « Bulliot, Bibracte et moi » en tant qu'ingénieur d'étude au CNRS rattaché au laboratoire Archéorient, à la Maison de l'Orient et de la Méditerranée.
 3. Archéorient, <<https://www.archeorient.mom.fr>>.
 4. Bibracte, <<https://www.bibracte.fr>>.

des beaux-arts et d'archéologie Joseph-Déchelette de Roanne⁵, la Société éduenne des lettres, sciences et arts de la ville d'Autun et la plateforme d'édition numérique Persée⁶.

Le projet s'organise en deux grandes phases. La première, qui s'est achevée en février 2021, concerne la transcription collaborative de ces archives associée avec des technologies de reconnaissance automatique d'écriture manuscrite. Notre article traite de cette étape sous la forme d'un retour d'expérience. La seconde phase se rapporte à leur publication sous forme d'une édition numérique enrichie sur la plateforme Persée d'ici la fin de 2022. Ces archives seront également intégrées à la médiation numérique mise en œuvre au Musée de Bibracte pour expliquer la chaîne opératoire de l'archéologie, dans une perspective historiographique concernant la découverte de la protohistoire européenne.

Nous décrirons en premier lieu les objectifs historiographique et épistémologique poursuivis par ce projet au travers de la réactualisation d'archives scientifiques du XIX^e siècle. Nous détaillerons ensuite sa méthodologie qui combine approches automatique et participative : le travail fastidieux se trouve confié à la machine, tandis que les participants se chargent du contrôle de la transcription et de son enrichissement. Enfin, nous exposerons les résultats de la phase de transcription et nous interrogerons les rapports que l'équipe a entretenus avec l'intelligence artificielle, qu'elle a contribué à former et qu'elle a manipulée tout au long du projet.

1. OBJECTIFS

1.1 Bibracte : de la guerre des Gaules à la civilisation des oppida

Le site de Bibracte se trouve en France, dans le massif du Morvan en Bourgogne, aux confins des départements de la Côte-d'Or, de la Nièvre, de la Saône-et-Loire et de l'Yonne. Bibracte était une ville fortifiée (*oppidum*) fondée à la fin du II^e siècle avant notre ère sur le mont Beuvray. Elle a été ceinturée de deux lignes de rempart, une de 7 km et l'autre de 5,2 km. Elle fut la capitale du peuple celte des Éduens durant un siècle, avant d'être délaissée au profit de la ville d'Autun. Ce lieu historique a été le théâtre de plusieurs épisodes marquants de la guerre des Gaules : César livra à proximité une bataille décisive contre les Helvètes en -58, Vercingétorix y

5. Musée de Roanne, <<http://www.museederoanne.fr>>.

6. Persée, « Bulliot », <<http://info.persee.fr/bulliot-bibracte-et-moi>>.

reçut en -52 le commandement de la coalition gauloise et Jules César s'y établit quelques mois pour achever le récit de ses campagnes militaires.

FIGURE 13.1
La Porte du Rebout



Principal accès à l'oppidum de Bibracte au I^{er} siècle avant notre ère.

Source: Wikimedia Commons.

Les premières fouilles débutèrent en 1864 sous l'impulsion d'Auguste Ernest d'Aboville (1819-1902), propriétaire des terrains, et de Xavier Garenne (1828-1883) dans un contexte marqué par la recherche des principaux sites mentionnés par César dans la *Guerre des Gaules* et la question de l'emplacement de l'ancienne ville de Bibracte. Napoléon III, qui a lancé des fouilles à Alise-Sainte-Reine (Alésia) dès 1861 et qui était en train de rédiger une histoire de Jules César, s'intéressa aux résultats de ces sondages. En 1867, il chargea Jacques-Gabriel Bulliot de diriger les fouilles du Beuvray et la Commission de la topographie des Gaules octroya une allocation annuelle pour leur financement. Les campagnes se succédèrent de 1867 à 1895. À partir de 1897 et jusqu'en 1901, Joseph Déchelette (1862-1914) poursuivit l'œuvre de son oncle. Une brève campagne en 1907 marqua la fin des fouilles jusqu'en 1984.

Le site de Bibracte permit à Joseph Déchelette de mettre en évidence la « civilisation des *oppida*⁷ », qui s'étendait, entre le III^e et le I^{er} siècle avant J.-C., du sud de l'Angleterre jusqu'en Europe de l'Est. Il s'est appuyé pour ce faire sur la similitude du mobilier archéologique mis au jour dans les *oppida* de Bibracte en France, de Manching en Allemagne, de Stradonice en République tchèque et de Velem-Szentvid en Hongrie. Interrompues depuis 1907, les recherches archéologiques reprirent sur le site de Bibracte à partir de 1984 avec la création du Centre archéologique européen. Cette structure assure chaque année l'accueil et le soutien logistique d'archéologues et d'étudiants issus de toute l'Europe et appartenant à une douzaine d'universités et d'instituts de recherche partenaires. Ses programmes de recherche portent aujourd'hui sur les modalités de l'urbanisation et de la romanisation.

1.2 Reconstituer le contexte archéologique des vestiges mis au jour au XIX^e siècle

L'archéologie a la particularité de détruire systématiquement le *contexte* de ses découvertes. Cette notion fondamentale désigne l'ensemble des informations associées au site et à ses vestiges : localisation, niveau archéologique et géologique d'origine, relations stratigraphiques entre les couches, associations entre les vestiges. Ces informations précisément recueillies constituent la documentation scientifique ou les archives de fouille. Elles représentent donc « une source incomparable d'information sur l'état de vestiges et de contextes désormais endommagés ou détruits » (Zanella *et al.*, 2017, p. 2).

Jacques-Gabriel Bulliot a fait preuve d'une « vision très moderne » (Guillaumet et Bonnefant, 1998, p. 6) de la documentation de fouille. S'il s'excuse du caractère « fastidieux et peut-être inutile » (Bulliot, 1899, t. 1, p. 116) qui pourrait transparaître de la « répétition des mêmes découvertes », d'une « nomenclature d'objets pareils dans des constructions uniformes » (Bulliot, 1899, t. 1, p. 116), il justifie ainsi de les enregistrer avec *intention* :

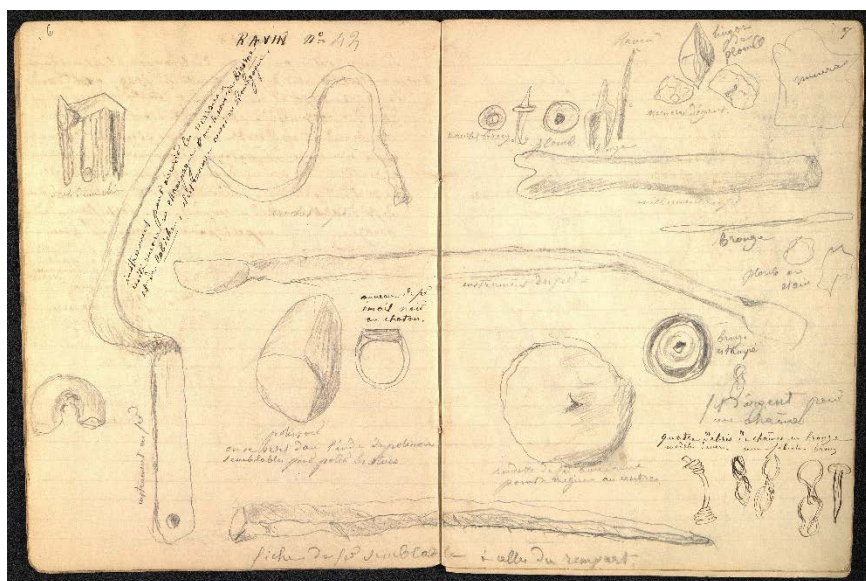
Ensevelies à la fin de chaque campagne sous le remblai, pour rendre le sol aux exploitants, les habitations de Bibracte ne laisseront d'autre souvenir que celui qui est consigné dans les procès-verbaux et les plans de fouilles, c'est pourquoi

7. Déchelette la caractérise notamment comme une civilisation urbaine, avec une spécialisation socioéconomique des quartiers, comme à Bibracte, des quartiers de métallurgistes et d'émailleurs et des quartiers résidentiels. Sur les discussions de cette notion, voir Philippe Gardes (dir.), « Oppida et "civilisation des oppida" cent ans après Joseph Déchelette », *Pallas, revue d'études antiques*, n° 105, 2017.

nous tenons à noter, à cataloguer, maison par maison, tous les détails qui permettent de restituer leur physionomie et leur destination (Bulliot, 1899, t. 1, p. 117).

L'importante documentation laissée par Bulliot permet d'avoir accès aux sources du savoir archéologique sur Bibracte. Les comptes rendus annuels qu'il publia dans les *Mémoires de la Société éduenne* ont été réédités en 1899 dans les *Fouilles du mont Beuvray (ancienne Bibracte), de 1867 à 1895* (2 vol.). Cette publication ne contient cependant pas de dessins ni de photographies du mobilier. Dans l'album de planches publié sous la direction de Félix et Noël Thiollier⁸, les principaux objets mis au jour sont présentés sans indication de provenance ni renvoi aux volumes de texte. Les carnets de fouille complètent ces publications imprimées en apportant des informations inédites, notamment sur le mobilier, qui a été dispersé entre différents établissements.

FIGURE 13.2
Les carnets de fouille de Jacques-Gabriel Bulliot



Ils comportent de nombreux dessins du mobilier mis au jour.

Source : carnet n° 4.1, 1870, p. 6-7.

8. Félix Thiollier et Noël Thiollier (dir.), *Fouilles du mont Beuvray, ancienne Bibracte. Album exécuté sous la direction de Félix et Noël Thiollier*, Saint-Étienne, impr. de J. Thomas, 1899, 61 pl., 1 plan.

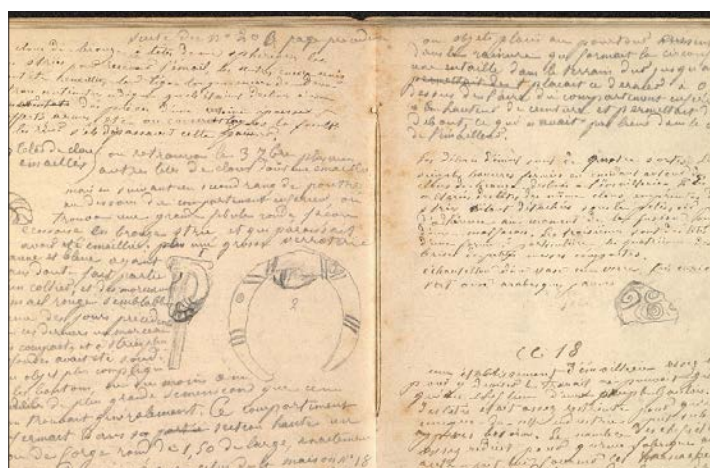
Dans le cas de reprise de sites qui, comme ceux de Bibracte et d'Alésia, ont déjà été fouillés au XIX^e siècle, le recours à ces archives permet de croiser données anciennes et contemporaines et d'établir le contexte de la découverte de certains vestiges archéologiques jusqu'alors sans provenance avérée et recueillis dans des conditions très éloignées des procédures scientifiques actuelles.

La consultation des archives de fouilles forme donc une étape essentielle pour la compréhension d'un site archéologique. Si elle est bien structurée, cette documentation historique⁹ fournira de nombreuses informations utiles à la reprise des fouilles, à leur orientation et à leurs interprétations. Certes, plus on remonte dans le temps et moins les descriptions et les méthodes de fouilles sont précises. Elles restent cependant une source complémentaire non négligeable. Dans un premier temps, on répertorie l'ensemble des travaux et ouvrages anciens publiés ou inédits ; dans un second temps on exploite les informations fiables en les intégrant dans le rapport de fouille. L'édition des archives de fouille permet de réactualiser des sources scientifiques anciennes. Ces documents inédits peuvent révéler des données oubliées et autoriser l'étude de collections d'objets à la lumière de sciences nouvelles (dendrochronologie, carpologie, archéozoologie etc.).

Il devient ainsi possible de comparer le mobilier décrit par Bulliot dans ses carnets, dont la découverte est datée et localisée, avec celui qui est dessiné sur les planches des albums de fouille conservés à la société éduenne d'Autun, et le mobilier dispersé entre plusieurs collections conservées au Musée d'archéologie nationale (Saint-Germain-en-Laye), au Musée Rolin (Autun) et au Musée de Bibracte (Saint-Léger-sous-Beuvray). Ces rapprochements permettront de cumuler et de corriger les informations contenues dans ces ressources documentaires (carnets manuscrits, inventaire des musées, publications imprimées, fonds iconographiques) et d'envisager des publications numériques enrichies, bien qu'un objet relevé sans stratigraphie perde de sa valeur scientifique.

9. Outre les archives manuscrites des archéologues, cette documentation, en grande partie inédite, est constituée de fonds de dessins, cartes et photographies anciennes, des mémoires des institutions savantes du XIX^e siècle.

FIGURE 13.3
Découverte d'une grande fibule ronde



La consultation des carnets permet de localiser et de dater la découverte d'objets répertoriés dans d'autres documents.

Sources : *Carnet n° 3*, 1869, p. 21, et *Fouilles du mont Beuvray*, album 57/1, s.d.

1.3 Des clefs de lecture pour l'archéologie du XIX^e siècle

Outre la connaissance de l'histoire particulière d'un site, les archives de fouille ouvrent des perspectives plus générales pour l'étude de l'histoire de l'archéologie et de ses méthodes. En ce qui concerne Bibracte, « l'évolution des problématiques et des techniques de fouilles mises en œuvre entre le milieu du XIX^e s. et la fin du XX^e s. fait du mont Beuvray un site prédisposé à l'histoire des sciences archéologiques » (Guillaumet et Bonnefant, 1998, p. 6).

On constate par exemple que Bulliot fut surtout attentif « à l'organisation spatiale du site, explorant par tranchées de vastes étendues mais négligeant le plus souvent l'approche stratigraphique, de sorte que [ses travaux reflètent] seulement l'état le plus récent des constructions dans chaque secteur » (Guillaumet et Bonnefant 1998, p. 7). Les ouvriers suivaient les murs et laissaient intact le centre des pièces. La fouille s'arrêtait généralement au niveau du premier sol rencontré, préservant ainsi les couches archéologiques plus profondes. Le mobilier récolté était composé de monnaies, d'objets métalliques et de céramiques décorées, et choisi selon des critères esthétiques. On retrouve aussi dans ses écrits les *a priori* de l'époque selon lesquels la civilisation ne peut provenir que de Rome.

En plus de ses techniques particulières, la compréhension de l'archéologie du XIX^e siècle suppose de mettre son vocabulaire en correspondance avec la terminologie actuelle. Ainsi une « poterie rouge fine » équivaut à de la sigillée¹⁰. Ce que Bulliot considère comme des sépultures est aujourd'hui interprété comme le remplissage de structures en creux pour divers usages : fosse, trou de poteau ou cave rebouchés par des déchets et des débris de construction. C'est la présence de mobilier et d'ossements ainsi que la couleur de la terre mêlée à des cendres et du charbon de bois qui ont fait croire à Bulliot que ces assemblages étaient des sépultures¹¹. Ce travail de rapprochement paraît essentiel pour l'indexation matière de tels corpus. Les toponymes sont également des ressources centrales pour l'indexation, l'agrégation et la recherche d'informations. Mais ils ne constituent pas des identifiants stables. Sur le mont Beuvray comme ailleurs, la dénomination et la localisation des secteurs de fouille ont évolué, tout comme celle des

10. Céramique fine caractéristique de l'Antiquité romaine, recouverte d'un engobe riche en oxyde de fer qui se vitrifie à la cuisson et prend une couleur rouge. Elle peut être ornée de décors en relief.

11. Nous devons ces indications à Jean-Paul Guillaumet qui, au cours de ce projet, a travaillé sur la terminologie employée par Bulliot et ses interprétations actuelles.

composantes de sites (quartiers, monuments, etc.). Une fois le texte établi et éditorialisé, l'exploitation scientifique des carnets Bulliot suppose un travail d'indexation préalable des mots clés et des toponymes. Ces référentiels de base serviront à la navigation dans le corpus et d'appui pour l'utilisation d'outils d'analyse textuelle et d'extraction des entités nommées¹².

2. MÉTHODOLOGIE : COMBINER INTELLIGENCE ARTIFICIELLE ET TRANSCRIPTION PARTICIPATIVE

2.1 Le corpus

Les carnets de fouille de Jacques-Gabriel Bulliot, conservés au musée Déchelette de Roanne¹³, forment un petit corpus de 864 feuillets, répartis en onze cahiers numérotés de 1 à 9, les cahiers 4 et 6 étant doubles. Ils couvrent une période discontinue, allant de 1867 à 1875, de 1880 à 1882 puis 1894. Ils ne présentent pas de cohérence chronologique ou topographique¹⁴. Ils n'ont jusqu'alors fait l'objet que de transcriptions partielles restées inédites, et réalisées sans protocole éditorial pour les recherches individuelles d'archéologues. Sans l'apport du collectif, des technologies numériques et de l'intelligence artificielle, ces archives seraient restées dans l'oubli car elles paraissent assez complexes à éditer.

La caractéristique de ce corpus manuscrit est d'être largement illustré de croquis et de plans, dans lesquels le texte, qui contient de nombreux chiffres avec les mesures des côtes, peut être positionné horizontalement ou verticalement. L'écriture de Jacques-Gabriel Bulliot évolue et devient tremblotante à partir de 1882 (carnet n° 9) et dans les pages du carnet n° 7 écrites en 1894.

-
12. Emmanuelle Perrin et Philippe Chassignet, « Extraire et structurer des informations géographiques sur les lieux de fouille archéologiques : le référentiel du site de Bibracte (mont Beuvray, Morvan) », *Assises MAGIS 2020*, GRD MAGIS, juin 2020, Paris.
 13. On peut consulter l'ensemble des carnets sur le site Mémo Roanne, « Fonds Joseph Déchelette. Bibracte Bulliot et Déchelette », <http://www.memo-roanne.fr/resultat.php?type_rech=fd&id_dam_collection=52&id_fonds=7&spec_expand=1&sort_define=score&sort_order=1&rows=9&fq=1>.
 14. De même, le recueil publié par Bulliot en 1899 propose, aux côtés de la table des matières classée par années, une table méthodique « pour suivre la description des fouilles dans l'ordre de chaque quartier ». Jacques-Gabriel Bulliot, *Fouilles du mont Beuvray (ancienne Bibracte)*, de 1867 à 1895, t. I, Autun, Dejussieu, 1899, p. 509.

fondée sur l'intelligence artificielle. Cette application a été développée entre 2016 et 2019 à l'Université d'Innsbruck (Autriche) dans le cadre du projet européen *Reconnaissance et enrichissement des documents d'archives (READ)*, avec un financement du programme de recherche européen Horizon 2020. Elle figure parmi les cinq lauréats 2020 du prix Impact horizon. Décerné par la Commission européenne, ce prix récompense des projets de recherche innovants à fortes répercussions sociales – et, en ce qui concerne Transkribus, l'accès aux archives et aux documents historiques. Transkribus est aujourd'hui une société coopérative européenne dont certains services sont payants. Le moteur de reconnaissance d'écriture manuscrite doit être configuré à partir de données d'apprentissage d'au moins 20 000 mots, obtenues par la transcription manuelle d'une centaine de pages. La technologie est fondée sur la segmentation ou la détection automatique des lignes qui permet d'associer l'image au texte et de restituer le contexte dans lequel les lettres sont employées.

2.3 Constitution et motivations de l'équipe de transcription

L'équipe du projet a été constituée lors d'ateliers de prise en main du logiciel Transkribus. L'atelier de lancement du projet a été organisé au Musée de Bibracte, à l'occasion des Journées du patrimoine, le 21 septembre 2019. Un deuxième atelier a eu lieu le 14 novembre 2019 au Musée Déchelette de Roanne, partenaire du projet. Une trentaine de personnes ont participé à chacun de ces ateliers. Après une démonstration individuelle de 15 à 20 minutes, l'application semblait rapidement maîtrisée et des groupes de travail de deux ou trois personnes se formaient pour transcrire un jeu de pages. Un troisième atelier de perfectionnement a rassemblé l'équipe en janvier 2020. Par souci de convivialité, chacune de ces rencontres comprenait également un déjeuner offert et pris en commun et une visite du site de Bibracte, du centre archéologique ou du Musée Déchelette en compagnie de responsables scientifiques.

Une dizaine de personnes ont composé de manière stable l'équipe de transcription. Elles décrivent leurs motivations par leur intérêt pour l'archéologie, pour l'histoire ancienne et locale, pour la généalogie et les archives, la paléographie et la transcription, voire par l'aspect informatique du projet, en ce qu'il pourrait s'appliquer aux documents d'archives à transcrire lors de leurs recherches généalogiques. Le renom de Bulliot et Déchelette apparaît encore comme un élément déterminant ajoutant une dimension scientifique à celle du patrimoine. La notion de plaisir semble

aussi essentielle. L'équipe se satisfait d'une œuvre utile, en lien avec l'écrit et la mémoire : transcrire c'est transmettre.

2.4 Organisation et déroulement du projet

Avec la transcription manuelle des trois premiers carnets, l'équipe a d'abord été chargée de la création des données pour la phase d'apprentissage supervisé. La création de ces données, associant l'image d'un mot à sa transcription, a impliqué de réaliser dans un premier temps une transcription semi-diplomatique, conforme au texte original de Bulliot. Le travail éditorial, avec les corrections orthographiques et l'harmonisation typographique, intervient dans la seconde phase de projet à la fin de 2022.

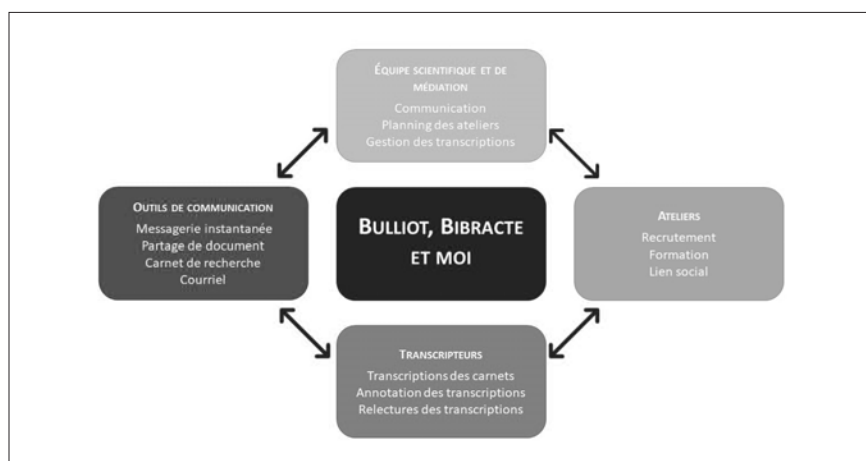
La coordination scientifique du projet était assurée par un ingénieur d'étude, spécialement recruté et chargé de la gestion des transcriptions : préparation (segmentation) et répartition des lots, suivies des transcriptions et des relectures. Chaque participant disposait d'un espace propre sur la plateforme Transkribus pour transcrire et annoter les jeux de pages qui lui avaient été attribués, les carnets ayant été divisés en 72 lots d'environ 12 pages chacun. Grâce aux droits d'administrateur, le coordonnateur gérait directement les transcriptions dans l'interface de Transkribus, en déposant les documents dans l'espace de travail de chaque participant et en archivant les travaux effectués selon les différentes étapes du travail. La transcription des carnets a demandé un important travail préparatoire de segmentation. La segmentation automatique a dû être entièrement revue pour numéroter les lignes selon la logique du texte. Les additions et corrections marginales ou interlinéaires, fort nombreuses, ont notamment été intégrées à leur place.

Le déroulement du projet a associé travail individuel et séances collectives. Neuf ateliers ont été organisés entre septembre 2019 et juin 2020, quatre en présentiel et cinq de manière virtuelle²⁰. Ces rencontres conviviales étaient l'occasion de résoudre en commun des difficultés de transcription et d'annotation et de progresser dans l'utilisation avancée de Transkribus (annotations, utilisation du moteur de reconnaissance d'écriture manuscrite, export d'une transcription au format texte). L'interface en

20. Sur les ateliers virtuels, voir Philippe Chassignet et Emmanuelle Perrin, « L'animation virtuelle du projet participatif Bulliot, Bibracte et moi », Webinaire Particip Act, « Participer depuis chez soi. Le confinement : source d'innovations pour les recherches culturelles participatives ? », 7 juillet 2020, <<https://www.youtube.com/watch?v=3G7y8l3tmTc&list=PLzek-91oARxuCwfWY-Q9rqVj11AP0bMYu&index=6&t=0s>>.

anglais de Transkribus a souvent constitué une difficulté que nous avons essayé de pallier en créant des tutoriels. Différents moyens de communication ont été mis en place pour fédérer l'équipe. Une messagerie instantanée, administrée par le coordinateur scientifique du projet, a permis de créer une dynamique d'échanges au quotidien entre tous les membres du projet. Un carnet de recherche dédié²¹ mettait à disposition divers documents utiles à la transcription, tels que des tutoriels sur Transkribus, une « foire aux questions », des guides de paléographie, des plans du mont Beuvray, des ressources onomastiques et bibliographiques. Il a également permis de valoriser les travaux de l'équipe de transcription avec la publication de « bonnes feuilles » extraites des textes transcrits.

FIGURE 13.5
L'organisation du projet « Bulliot, Bibracte et moi »



Source : Schéma P. Chassignet.

L'organisation du projet tenait à éviter toute verticalité dans le dispositif de vérification des transcriptions (voir la figure 13.5). Le processus de validation a donc impliqué l'ensemble des participants avec l'organisation de relectures croisées entre pairs et la recherche d'un consensus par la discussion, lors d'ateliers, sur les mots difficiles à déchiffrer. Un protocole a été mis en place pour établir les termes dont la lecture restait incertaine, en recherchant ou en vérifiant l'occurrence d'un toponyme, d'un patronyme,

21. Bulliot, Bibracte et moi, « Transkribus au programme de deux rencontres scientifiques sur la transcription automatique », <<https://bbm.hypotheses.org/>>.

d'un mot ou d'une expression dans les publications imprimées de Bulliot et dans les dictionnaires du XIX^e siècle.

À la différence d'autres projets participatifs, le travail fastidieux n'était pas confié aux bénévoles, mais au coordinateur du projet pour ce qui concerne la vérification et la correction de la segmentation automatique des pages, et à l'intelligence artificielle avec la transcription automatique. Aux membres du projet étaient dévolues les tâches habituellement réservées aux spécialistes : la validation et l'enrichissement de la transcription. Envers la machine, ils étaient ainsi placés dans un rapport d'apprentissage – ils lui apprenaient à lire – et de contrôle, avec la relecture et la correction des transcriptions automatiques. L'intelligence artificielle a donc eu le rôle d'un *assistant à la transcription*, qui aide l'équipe dans le déchiffrement des manuscrits.

3. RÉSULTATS

3.1 La configuration d'un modèle pour la transcription des manuscrits français du XIX^e siècle

Un des résultats de la première phase du projet « Bulliot, Bibracte et moi » est la configuration et la mise à disposition d'un modèle pour la transcription de manuscrits français du XIX^e siècle. Un premier moteur de reconnaissance de l'écriture de Jacques-Gabriel Bulliot a été configuré en mars 2020 à partir des trois premiers carnets de fouille (24 500 mots). Un travail de nettoyage avait été réalisé en amont par le coordonnateur du projet pour aplanir les difficultés que pouvait rencontrer l'intelligence artificielle, avec la vérification de la segmentation, la suppression des chiffres, des mots trop raturés et des lignes verticales. Grâce à ce travail de nettoyage et aux doubles relectures assurées par l'équipe de transcription, le taux d'erreur par caractère sur le jeu de données d'entraînement est passé de 7,54 % à 3,46 % et le taux sur le jeu de données de validation²², de 9,38 % à 6,11 % par rapport à un premier moteur test configuré sur les carnets non relus. Un second moteur configuré sur l'ensemble des carnets (147 000 mots) a été mis à la disposition sur le site de Transkribus en mars 2021²³. Il fait partie des modèles publics que chacun peut utiliser pour transcrire

22. Le jeu de données de validation est un lot qui n'est pas intégré aux données d'apprentissage et qui va servir à tester le moteur.

23. « French handwriting 19th century free public AI model for handwritten text recognition with Transkribus », <<https://readcoop.eu/model/french-handwriting-19th-century>>.

des documents similaires, c'est-à-dire des manuscrits du XIX^e siècle en français, et continuer à améliorer. Le taux d'erreur est assez élevé (7,73 %) toujours en raison des plans et des dessins dont les cotes et les légendes sont orientées de manière hétérogène. Un travail de nettoyage comme celui de la configuration du premier moteur aurait sans doute amélioré ces performances.

FIGURE 13.6

**Comparaison de la transcription automatique (en rouge)
et de la transcription humaine (en vert) dans le jeu de données
de validation du modèle définitif du projet**

1-1 # Fouilles du Beuvray
1-2 # ~~186m~~ 1867
1-3 # Champlain-champlain
1-4 # maison-Maison de la butte du Champlain-champlain
1-5 # n°2-n° 2
1-6 # à dix mètres-mètres au sud de la crête du
1-7 # retranchement supérieur, ou au revers nord de
1-8 # la butte du Champlain-champlain, vaste construction avec
1-9 # assises en pierres brutes-brutes reliées par la terre glaise, au-glaise, aire travail rudimentaire-rudimentaire
1-10 # enfouie à 2,30 1,30 de 7-refondreur-profondeur.
1-11 # et peu solide-solide, mur du Dordodéteur-nord détruit au pular-milieu.
1-12 # Elle a enlouve 7m15 en l'oeuvre 7.m15
1-13 # du Nord-nord au sud et 7m 7. m de l'est à l'ouest.
1-14 # L'aire, enfouie à 1m30 1.m30 de profondeur est en terre
1-15 # glaise battue x
1-16 # --x 0,30 d'épaisseur d'épaisseur
1-17 # det-repose et repose sur une couche de
1-18 # cailloutis lave-lavé d'un mètre d'épaisseur d'épaisseur,
1-19 # mais au milieu de la pièce, la terre glaise
1-20 # durcie par le feu est passées-passée
1-21 # à l'état de frique-brique
1-22 # et formé-forme, sur un
1-23 # à deux mètres carrés
1-24 # 060.9 m.
1-25 # 040 aire brûlée
1-26 # un véritable carrelage-carrelage en tere-cuite terre cuite.
1-27 # le-ce point devaitêtre-devait être le foyer-le foyer, à un mètre
1-28 # du mur sur-sud un petit compartiment carré
1-29 # de 0,30 sur 0,20 de côté, dont le fond
1-30 # est pacée pavé d'une tuile et faut-dont les bords
1-31 # sont en-constrits-circonsrits par quatre pierres placées
1-32 # sur champ, épaisses de trois à quatre centimètres-centimètres,
1-33 # estenlouie-est enfouie à 0,10 au dessous de l'aire,
1-34 # la tuile est placée sur un cailloutis-batte cailloutis battu avec de

3.2 La préparation de l'édition scientifique des carnets de fouilles de Bulliot

Cette première phase a également permis de préparer l'édition numérique des carnets selon les conventions éditoriales des éditions scientifiques. Il s'agit par exemple de l'établissement des notes sur le manuscrit ou notes critiques. Les additions et corrections marginales ou interlinéaires sont intégrées à leur place dans l'édition, mais leur présence et leur position sont signalées en note, ainsi que les mots raturés, dans la mesure où ils restent lisibles. L'équipe de transcription a contribué à ce

type d'annotation en employant les balises proposées sur la plateforme Transkribus et fondées sur le standard TEI. Ont ainsi été indiqués les incertitudes de lecture (*unclear*), les mots illisibles (*gap*), les lettres restituées en raison d'une lacune physique (*supplied*), les ajouts dans les marges ou entre les lignes (*add*). Les abréviations ont été signalées et développées (*abbrev*). Un balisage sémantique a également été réalisé pour indexer les entités nommées ainsi que les dates et les références bibliographiques : noms de personnes (*persName*), noms de lieux (*placeName*), indications de date (*date*), références bibliographiques (*work*).

Pour le suivi des relectures, les commentaires permettaient de laisser à propos du choix d'une transcription des informations qu'aucune balise ne pouvait signaler. Nous avons également convenu d'utiliser les commentaires pour indiquer les textes biffés, les mots sur lesquels un autre mot a été réécrit et un passage au crayon sur une page écrite à l'encre et inversement.

À l'issue de cette étape de transcription et d'annotation, les carnets ont été remis à Persée pour la seconde phase en cours du projet, sous différents formats (XML ALTO, pdf interrogeable et tableur pour les balises). Ils sont en cours de structuration et d'indexation pour être mis en ligne dans le cadre d'un site dédié (Perséides) par cette unité d'appui et de recherche.

3.3 Le compagnonnage de Wolpertinger

Quels rapports les membres du projet « Bulliot, Bibracte et moi » ont-ils entretenus avec l'intelligence artificielle et avec le modèle de reconnaissance d'écriture manuscrite, qu'ils ont contribué à former et qu'ils ont manipulé ?

Dans les échanges, ce dernier est désigné sous le nom du logiciel, de moteur, d'automate ou surnommé Wolpertinger, du nom de la chimère qui sert de logo à Transkribus. Cet animal légendaire des forêts des Alpes bavaoises est souvent représenté avec des ailes, des bois de chevreuil, des crocs et une crête de coq.

Lors de leur première utilisation de la transcription automatique, en mars 2020, dans les commentaires échangés sur messagerie instantanée, les membres de l'équipe se disent impressionnés par sa fluidité et sa rapidité lorsque le texte est régulier : « Il y a des erreurs mais je suis quand même impressionnée par le résultat. » « 5^e page long texte, impressionnant ! » « Le

moteur travaille bien voire même très bien lorsqu'on est en pur texte». «J'ai fait quatre pages. Je suis bluffée par la transcription²⁴».

FIGURE 13.7
Un wolpertinger



Une représentation de l'intelligence artificielle? Un wolpertinger composé avec des ailes, des crocs et des bois à partir du *Lièvre*, d'Albrecht Dürer

Source : Wikimedia Commons.

L'équipe constate aussi que l'intelligence artificielle «bute là où on aurait buté»: «Il semble que le pauvre Transkribus rencontre les mêmes difficultés que nous à différencier les «a» des «o» et les «n» des «u» de Bulliot.» «J'ai presque terminé la relecture de la première page du lot automatiquement retranscrit. Sur cette page j'estime à environ 10 % les erreurs rencontrées, et pour la plupart sur des endroits où j'aurais eu les mêmes difficultés à retranscrire une écriture rapide, où l'on devine plus les mots par le contexte qu'on les lit réellement. L'automate fonctionne bien.» «Souvent les erreurs de Transkribus correspondent à des zones où c'est

24. Toutes les citations sont extraites de la messagerie instantanée du projet.

notre connaissance du contexte ou de la langue française qui nous permet de donner la bonne lecture.»

Les faiblesses récurrentes du modèle avec les croquis sont rapidement repérées : « Les choses se gâtent pour les chiffres, et surtout les zones incluant du dessin. » La transcription est « fluide pour le texte, plus chaotique pour les chiffres, les majuscules... » Les cotes des plans « inscrites dans différents sens n'ont pas été transcrites ou de manière erronée voire farfelue ». Autre obstacle, l'écriture changeante et tremblante d'un Bulliot vieillissant qui « fait aussi trembler Transkribus » : « Wolterpinger se déchaine ou est complètement épuisé. »

Les pages apparaissent beaucoup plus rapides à relire, « quand on a bien compris les manies orthographiques de Wolpertinger, avec quelques succulents néologismes ! » Amusés par ces erreurs de transcription ou par les « fantaisies de Wolpertinger », des membres de l'équipe en relèvent quelques-unes pour les partager avec le groupe : « le moteur a complètement pataugé avec *"ciuen e le aule peler de bour"* alors que je transcris sans soucis *"Maison en bois en avant du péristyle du temple"*. » « Un peu plus de matière pour enrichir notre lexique de patois HTR : *"daule de 11 5. dute la gaconde hauvaur cheraux"* à traduire aisément par *"Fouilles de 1872- Suite de la maison du parc aux chevaux"*. » Un peu plus loin sur la même page : « *étrabandonnée*, là, n'est-ce pas une transcription de Claude Ponti ? »

L'équipe constate que la correction des transcriptions automatiques implique une relecture lettre à lettre très différente de l'opération de déchiffrage que demande une transcription manuelle. Elle oriente également la lecture de certains mots. Aussi certains membres préfèrent s'en détacher et lire le texte manuscrit avant de voir la transcription automatique. On relève aussi des critiques concernant la fatigue de travail de relecture sur écran, où les fautes semblent plus difficiles à voir que sur le papier. L'équipe reste aussi assez circonspecte quand il lui semble que « l'homme travaille pour la machine » plutôt que l'inverse. C'est le cas face aux difficultés de la transcription automatique de l'écriture vieillissante de Bulliot – les données étaient insuffisantes pour parfaire l'apprentissage de cette « seconde main » – ou quand le travail de transcription doit être entièrement repris à la suite d'erreurs de manipulation. Le logiciel paraît donc adapté et efficace surtout pour la transcription de documents réguliers et répétitifs comme le sont les actes notariés et administratifs.

En conclusion de ce retour d'expérience sur la première phase du projet « Bulliot, Bibracte et moi », on constate que l'exploitation de ces documents

transcrits grâce à l'intelligence artificielle associée au participatif devrait conduire à de nouvelles expérimentations dans ce domaine. En plus des techniques d'analyse textuelle et d'extraction des entités nommées que nous avons déjà évoquées, la reconnaissance automatique des images aiderait par exemple à associer les croquis du matériel archéologique consignés dans les carnets de fouilles avec les bases de données des musées où sont conservés les mêmes objets. Ces rapprochements iconographiques permettront de cumuler et de corriger les informations contenues dans ces ressources documentaires et de restituer les contextes archéologiques d'objets aujourd'hui encore privés de provenance et d'identification précises. La gestion documentaire spécifique des croquis, qui apparaît comme un problème récurrent avec ce type de document, appelle à des développements technologiques intéressants, tant au niveau sémantique, avec leur légendage et la génération automatique de descripteurs textuels, qu'au niveau du lien entre les données avec le traitement des renvois entre un texte et ses illustrations.

La documentation ancienne est nécessaire aux fouilles d'aujourd'hui. On l'a vu, l'essentiel des informations sur les objets mis au jour au XIX^e siècle sont encore inédites et ne se trouvent que dans les carnets de Bulliot. Face à la « documentation archéologique immergée²⁵ » que constituent ces fonds d'archives, l'association de la science participative et de l'intelligence artificielle montre sa force et son efficacité pour faire œuvre de transmission.

RÉFÉRENCES

- Aubert, Olivier, Guillaume Raschia et Benjamin Hervy, « IA et approches participatives dans les humanités numériques : de la conjonction à la coopération », dans Martin Grandjean et al. (dir.), *Colloque Humanistica 2021 : recueil des résumés*, 2021, p. 32-33.
- Bulliot, Jacques-Gabriel, *Fouilles du mont Beuvray (ancienne Bibracte), de 1867 à 1895*, Tome I, Autun, Dejussieu, 1899, 2 vol.
- Hecker, Suzanne et al. (dir.), *Citizen Science. Innovation in Open Science, Society and Policy*, London, UCL Press, 2018.
- Gardes, Philippe (dir.), « Oppida et "civilisation des oppida" cent ans après Joseph Déchelette », *Pallas, revue d'études antiques*, n° 105, 2017.

25. Pour reprendre l'heureuse formule du projet Cardo : « Carnets de fouilles. La documentation archéologique immergée (1898-1986). Deux cas d'étude, entre Gaule et Italie préromaine ». Voir *Les passés dans le présent*, <<http://passes-present.eu/fr/carnets-de-fouilles-la-documentation-archeologique-immeree-1898-1946-deux-cas-detudes-entre-gaule>>.

- Gefen, Alexandre, « Les promesses de l'intelligence artificielle pour les humanités numériques », conférence inaugurale du colloque *Cahier*, « 10 ans de corpus d'auteur », 7 juin 2021.
- Gruel, Katherine et Daniele Vitali (dir), « L'Oppidum de Bibracte. Un bilan de onze années de recherche (1984-1995) », *Gallia*, n° 55, 1998.
- Guillaumet, Jean-Paul et Pierre-Paul Bonenfant, « Les recherches de J.-G. Bulliot et J. Déchelette à la lumière des fouilles modernes », *Gallia*, « L'Oppidum de Bibracte. Un bilan de onze années de recherche (1984-1995) », n° 55, 1998, p. 6-8.
- Hervy, Benjamin *et al.*, « Correction des données : retour d'expérience sur la plate-forme RECITAL de transcription participative », *Recherche d'information, document et Web sémantique*, vol. 18-2, 2018.
- Houllier, François, *Les Sciences participatives en France : état des lieux, bonnes pratiques et recommandations*, Rapport élaboré à la demande des ministres en charge de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, 2016.
- Muehlberger, Guenter *et al.*, « Transforming scholarship in the archives through handwritten text recognition : Transkribus as a case study », *Journal of Documentation*, vol. 75, n° 5, 2019, p. 954-976.
- Perrin, Emmanuelle et Philippe Chassignet, « Rendre visible la face cachée de l'iceberg. Explorer les documents manuscrits à l'aide de l'intelligence artificielle », *ArchéOrient – Le Blog*, 27 mars 2020, <<https://archeorient.hypotheses.org/14807>>.
- Zacklad, Manuel et Lisa Chupin, « Le crowdsourcing scientifique et patrimonial à la croisée de modèles de coordination et de coopération hétérogènes : le cas des herbiers numérisés », *La Revue canadienne des sciences de l'information et de bibliothéconomie*, vol. 39, n°s 3-4, 2015, p. 308-328.
- Zanella, Sandra, Jean-Pierre Brun, Martine Denoyelle, Pierre Rouillard et Stéphane Verger (dir.), *Les Archives de fouille : mode d'emploi*, Paris, Collège de France, 2017.

INTRODUCTION DE L'INTELLIGENCE ARTIFICIELLE DANS LE MÉTIER DE GUIDE

QUEL IMPACT SUR LA PRATIQUE DE LA VISITE GUIDÉE ?

Lina Uzlyte¹

Le métier de guide touristique, art de l'accompagnement, est aussi un média constitué autour de la pratique culturelle qu'est la visite guidée. Un média est un rapport structuré par une technique et par des logiques économiques et professionnelles (Boucher, Duquet et Beaudry, 2008). Il s'instaure entre le producteur des contenus et son public par la séduction, la confiance, mais aussi la dépendance. Par exemple : les techniques narratives, l'usage du direct, d'un langage simplifié, l'importance accordée au cadre de l'action. Ces éléments définissent les contours de la mission de guide qualifié consistant à faciliter la compréhension du lieu visité. La visite guidée, résultat de l'action de guide, est au centre de son activité professionnelle. Quand il s'agit de guide qualifié, la visite guidée représente la pratique culturelle de la médiation patrimoniale – processus d'information et de communication. Ces notions peuvent être reliées par une théorie de l'informatique qu'on nomme intelligence artificielle² (IA).

-
1. Chercheuse postdoctorante en sciences de l'information et de la communication à l'Université de Reims.
 2. Selon l'*Oxford English Dictionary*, l'IA est une théorie et un développement de systèmes informatiques capables d'effectuer des tâches nécessitant une intelligence humaine, telles que perception visuelle, reconnaissance de la parole, prise de décision et traduction des langues. Merriam-Webster décrit l'IA comme une section de l'informatique traitant de la simulation

À partir de l'introduction de l'intelligence artificielle dans la pratique de la visite guidée qu'assure le guide qualifié, sa compétence interculturelle se voit soulignée. Face à la crise sanitaire, la mobilité des individus s'est vue brutalement limitée et la fermeture des lieux patrimoniaux a obligé à repenser leur fonctionnement habituel. De nombreux dispositifs numériques ont été récemment proposés pour accompagner les visiteurs sur les territoires visités. Cependant, cette pratique impose des procédés qui remettent en question l'accès à la culture et ont tendance à renforcer les inégalités (Rouvroy et Berns, 2013). Cette étude se propose, dans le contexte européen, d'observer comment l'introduction de l'IA réoriente la pratique de guide qualifié aujourd'hui.

Dans un premier temps, nous verrons comment l'IA investit les espaces de la visite guidée. Dans un deuxième temps, les évolutions constatées seront mises en parallèle avec le code déontologique de la profession, en dégagant les valeurs et les principes qui stimulent précisément cette évolution.

Cette étude puise les données d'un travail doctoral plus vaste sur la conception, l'organisation et l'exercice de l'activité de guides en Europe entre 1990 et 2020 (Uzlyte, 2021). Des entretiens ont été menés auprès des guides experts sur la libre circulation des personnes, qui à son tour ouvrait à des interrogations relatives à l'harmonisation, à la qualification et à la compétence particulière de la profession en Europe. Ce travail a permis d'examiner la construction de la définition de l'activité de guide et sera utilisé ici comme le point de départ.

La définition actuelle de l'activité de guide, qui émerge des chartes établies par le Conseil international des monuments et des sites (ICOMOS), exprime la voix des experts en ce domaine³. Ainsi, elle devient normalisée et apparaît d'abord comme un consensus entre gouvernements européens et professionnels du secteur. Dans ce cadre, la fonction de guide

du comportement humain. *L'Encyclopaedia Universalis* caractérise l'IA comme un stimulus des différentes facultés cognitives constitutives de l'intelligence, qu'elle soit humaine ou animale, sur des ordinateurs, en les programmant avec les techniques du traitement de l'information.

3. Voir la Charte de Venise qui lance le mouvement évoqué en 1964. Nous ne retenons ensuite que celles qui mettent l'accent sur la gestion des interventions sur les sites, territoire privilégié des guides : chartes relatives au tourisme culturel (1976, révisée en 1999), aux jardins historiques (1982), aux villes historiques (1987), à la gestion du patrimoine archéologique (1990), au patrimoine culturel subaquatique (1996), aux itinéraires culturels (2008), à l'interprétation et à la présentation des sites culturels patrimoniaux (2008), à la sauvegarde et à la gestion des villes et ensembles urbains historiques (2011), enfin, aux paysages ruraux (2017).

est caractérisée par la notion d'interprétation du patrimoine, notion construite par les structures politiques. Le contenu de ces documents permet de constater qu'ils l'abordent selon trois dimensions : 1) *instrumentale* ou *méthodologique*, soit les qualifications et la formation standardisée des professionnels du patrimoine ; 2) *éthique* ou *sociale*, soit la relation entre sites et visiteurs et visiteuses lors de la visite guidée ; 3) *théorique* et *philosophique*, c'est-à-dire la manière de comprendre le patrimoine, sa sauvegarde et le contexte particulier du guidage.

L'enquête menée auprès des guides experts, mais aussi auprès des concepteurs et conceptrices de l'activité institutionnels, a rendu les données présentées plus loin opérationnelles au moyen des actions suivantes :

- Un questionnaire exploratoire établi lors des réunions professionnelles internationales de l'European Federation of Tourist Guide Associations (FEG) et les guides de la région de la mer Baltique.
- Des entretiens semi-directifs sur la pratique professionnelle de guide européen (conception, organisation, exercice de l'activité en Europe) vue par les guides experts et expertes eux-mêmes.
- Des entretiens compréhensifs semi-directifs auprès des autorités responsables de la conception de la profession.
- L'analyse des réunions interministérielles autour de la profession de guide-conférencier en France, de 2015 à 2018, afin d'aborder le point de vue institutionnel.
- L'analyse des commentaires accompagnant la pétition⁴ des guides-conférenciers et des guides-conférencières français ;
- Un questionnaire sur les attentes des visiteurs et des visiteuses, dans les pays baltes.

Enfin, ces données ont été complétées par le corpus documentaire portant sur la définition et l'usage de l'IA et les questions qu'elle soulève dans le quotidien de guide qualifié. L'approche qualitative du terrain a été privilégiée pour la particularité des données qualitatives correspondant aux situations de guidage, ce qui nous a permis de nous tenir au plus près de l'expérience vécue par les guides.

4. Voir la pétition du 11 mai 2015 disponible en ligne. Mes opinions, « Soutien aux guides-conférenciers », <<https://www.mesopinions.com/petition/art-culture/soutien-aux-guides-conferenciers/13370>>, (consulté en février 2022).

1. LE MÉTIER DE GUIDE ET L'INTELLIGENCE ARTIFICIELLE

Il ressort des rapports nationaux rendus à la Commission européenne en 2015 la volonté de réorganiser le cadre de la profession dans la totalité des pays européens. D'après le référentiel-métier européen, le guide est employé en vue d'accompagner les visiteurs et les visiteuses en interprétant les lieux visités⁵. Cette compétence particulière est décrite comme un processus de communication ayant pour but de révéler la signification du lieu ou de l'objet visité⁶. Les processus de communication autour du patrimoine se forment, aujourd'hui, sous l'influence des techniques de l'information et de la communication, lesquelles réservent une place importante à l'interprétation des contenus et à leur sens. Le fait de produire un récit autour des mêmes objets abordés, soit les positions particulières des acteurs, dans un cadre stabilisé, transforme ces lieux en espace de communication spécifique – espace spécifié d'abord par le discours issu des politiques publiques, puis par le champ d'action et orienté, enfin, par la nature de l'activité. Des dispositifs à la fois stratégiques, juridiques et économiques, ainsi que les offres éducatives se sont alors structurés. En ce sens, l'interprétation engage une communication propre au domaine de la médiation culturelle. Ainsi, nous comprenons la compétence particulière de l'interprétation comme une approche communicationnelle du territoire, au sens où l'entend Michèle Gellereau : « [...] le guide, dans son travail d'interprétation, est un acteur qui met en lumière un caractère important de la médiation culturelle : une rencontre participant à la construction d'un "monde commun", dans des modalités diversifiées où l'interprétation des lieux se construit à plusieurs » (Gellereau, 2004).

De plus, les guides reçoivent en héritage⁷ une catégorisation établie en fonction des territoires couverts par leurs compétences spécifiques, en

-
5. Il s'agit de la Classification européenne des aptitudes/compétences, certifications et professions (ESCO). Elle est pertinente pour le marché du travail, l'éducation et la formation dans l'Union européenne (UE). La hiérarchie des professions à l'ESCO est issue de l'utilisation de l'ISCO-O8 (OIT). Les guides assistent les individus ou les groupes lors d'une installation artistique, d'un voyage ou d'une visite dans des lieux d'intérêt touristique, tels les musées, les monuments et les lieux publics. Par les informations qu'ils fournissent, ils leur permettent d'avoir accès à l'interprétation du patrimoine culturel lié à un objet, à un lieu ou à une région.
 6. Note l'Association for Heritage Interpretation (AHI), « The professional network for heritage interpreters », <<https://ahi.org.uk>>, (consulté en février 2022).
 7. Voir *Dictionary of Occupational Titles, Index revised, 4th edition*, fourni, à titre de service public, par Information Technology Associates (ITA). Obsolète aujourd'hui, car il est considéré par les autorités comme trop complexe pour l'utilisation automatique. Ce référentiel classe les guides dans l'ordre suivant : « guide », « guide de l'aéroport », « guide de croisières », « guide

matière d'interprétation du patrimoine⁸ (Lingle Pond, 1993, p. 30, 168). Depuis 2003, les guides européens « interprètent le patrimoine culturel et naturel d'une aire géographique déterminée », sur la base d'une « qualification spécifique pour guider sur ce territoire⁹ ». Ce faisant, ils participent à l'élaboration de la politique publique de la conduite de visites guidées dans les espaces publics. Ainsi, à travers la relation des guides à leur champ d'action, nous retrouvons la caractéristique précise de leur action. Il importe ici de noter qu'« en vingt ans, nous sommes passés d'une résistance aux écrans à un désir d'écrans dans l'exposition au niveau des discours institutionnels » (Schmitt et Meyer-Chemenska, 2015). Alors que dans les années 1990 les musées et le secteur du patrimoine témoignaient d'une certaine méfiance envers la numérisation, force est de constater que l'État s'affirme, en 2000, comme un opérateur incontournable de la numérisation du patrimoine et que le numérique devient un enjeu majeur de la politique culturelle (Eidelman, 2017).

Dès lors, experts et expertes sont sollicités afin de mener les campagnes analytiques sur les effets du numérique, avec, souvent, le soutien financier de la direction régionale et le soutien à la fois financier et technologique de l'Action européenne.

C'est ainsi que la panoplie d'instruments d'investigation et de diagnostic mobilisés pour l'acquisition de connaissances préalables s'enrichit de capteurs et d'algorithmes d'analyse et de traitement de l'image. [...] Les responsables de sites et de collections font désormais appel à des sociétés prestataires de services maîtrisant les technologies de numérisation, de traitement et de visualisation interactive, afin d'élaborer des circuits de

-
- délégué», «guide accompagnateur», «guide des publics locaux», «guide des publics étrangers», «guide de pêche ou de chasse», «guide des plantes», «guide des sites, cicerone».
8. Kathleen Lingle Pond spécifie la catégorisation suivante : guides touristiques (*tour guides*), guides des villes (*urban guides*), guides officiels (*Government guides*), guides-conducteurs (*driver-guides*), guides des industries (*business guides or industry guides*), guides de loisirs (*adventure guides*), guides accompagnateurs (*tour managers*), guides interprètes (*interpreters*) et guides des musées : bénévoles (*docents : volunteer guides*).
 9. Voir la norme EN 13809:2003 du Comité européen de normalisation (CEN) qui fixe cette obligation. La Directive 2013/55/UE porte, elle, sur la transparence et l'évaluation des professions réglementées ; elle définit les règles de la reconnaissance professionnelle en précisant l'équivalence des qualifications professionnelles, les conditions de la libre prestation de services et d'établissement de l'activité au sein de l'Union européenne. Voir CE, Directive 2013/55/UE du Parlement européen et du Conseil du 20 novembre 2013 modifiant la directive 2005/36/CE relative à la reconnaissance des qualifications professionnelles et le règlement (UE) n° 1024/2012 concernant la coopération administrative par l'intermédiaire du système d'information du marché intérieur («règlement IMI»), [2013] JO, L 354/132.

visite, des préparations à la visite, des reconstitutions/restitutions d'œuvres, d'objets, de sites, des dispositifs interactifs, etc. (Liévaux et De Luca, 2020).

Ces recherches donnant à voir le paysage des applications de l'IA, qui est constitué des techniques d'analyse non invasives (FRX1) et micro-destructives (LIBS2), combinées à différentes techniques d'imagerie (macro et microphotographie technique dans le visible (VIS), l'ultraviolet (UV) et l'infrarouge (IR), photogrammétrie, imagerie hyperspectrale ont permis de déterminer le niveau d'authenticité des œuvres exposées et d'évaluer leur état de conservation (Degrigny, Farrugia, Mérienne et Pinette, 2020). La retranscription visuelle des informations collectées, au moyen de la réalité augmentée, est intégrée désormais dans le parcours de la visite des musées et des sites.

L'imagerie numérique s'est imposée d'une part comme un moyen de médiation offrant une plus-value scientifique et pédagogique à la visite (Courrèges d'Agnos, Dubos et Legrosse, 2020). L'objectif du recours à la 3D est multiple : il s'agit de numériser les œuvres difficilement accessibles – parfois même avec des parties non visibles, donc difficiles à apprécier pour le visiteur – et de permettre leur restitution par l'infographie ou par des dispositifs interactifs (Boudon-Machuel et Serres, 2020). L'ensemble des techniques et du matériel permettant de mélanger de façon cohérente des éléments réels et virtuels, au sein d'une même scène, est désigné par les termes de réalité augmentée.

D'autre part, la massification des pratiques numériques et la généralisation des pratiques de surveillance ont contribué au fait que, par le biais des algorithmes, les mondes numériques sont devenus un système de contrôle des conduites (Cardon, 2017). « Les rétentions tertiaires numériques des données personnelles qui deviennent de moins en moins personnelles, de moins en moins secrètes et intimes, de plus en plus standardisées par des programmes et les réseaux sociaux, contribuent à la construction de *big data* » (Stiegler, 2021). Déterminer la conduite d'un individu ou d'un groupe d'individus en fonction de stratégies en utilisant des tactiques révèle la notion de gouvernementalité, proposée et décrite en amont par Michel Foucault. Selon l'auteur : « Lorsqu'on est gouverné, on accepte aussi d'être gouverné. Et donc les conditions de notre obéissance au pouvoir sont toujours internalisées ; elles sont toujours accordées à la volonté de l'individu » (Foucault, 1983). De ce fait, il est possible d'évoquer une forme de gouvernementalité par l'environnement déployé par l'architecture algorithmique (Stiegler, 2017), où l'individu représente

un espace potentiel qui sera soit activé, soit annulé, selon sa manière de consommer les produits numériques. Il est alors intéressant d'observer comment et pourquoi le rapport au numérique fait problème et sous quelle forme il modifie la nature des savoir-vivre, savoir-faire, savoir-agir et savoirs conceptuels des guides qualifiés.

1.1 Principes et valeurs émergeant de la mise en application de l'IA à la situation de guide

Aujourd'hui, les « 23 principes d'Asilomar » fondent le principe de précaution, qui devrait être appliqué à l'IA sans empêcher la connaissance de progresser et pour maîtriser les risques. Ce principe de précaution repose sur des valeurs de sécurité, transparence, responsabilité, liberté et respect de la vie privée, intérêt général et contrôle¹⁰. Juridiquement, le principe de précaution relève de la responsabilité des États. Les États signataires ont l'obligation d'intervenir dès que la sécurité pour l'environnement ou la santé humaine ou animale est menacée. Ainsi, ce principe a été adopté dans le Traité de Maastricht et enfin nous avons les règles déterminées par le Sénat français afin d'accompagner les usages de l'IA par une réflexion éthique (De Ganay et Gillot, 2017). Cela se traduit par les principes suivants : positionnement d'un discours professionnel clair et transparent, communication à destination des publics mesurée et pédagogique, approche éclairée et éclairante du public (sensibilisation à ces principes), protection des données personnelles et contrôle de sa conformité à la régulation, vigilance contre la fraude numérique systématisée et organisée, respect et prise en compte des situations « sensibles » liées aux statuts sociaux et professionnels des intercalants, réflexion critique sur l'activité et le système de communication (le code déontologique élaboré par un collectif des professionnels peut suffire pour les déclarer en tant que profession). Dans la situation décrite, définir la numérisation devient nécessaire. L'Association des professionnels de l'information et de la documentation (ADBS) propose de la comprendre ainsi : « processus de conversion d'un signal analogique en un signal numérique. Cette technique permet de stocker des documents, quels qu'en soient la nature ou le support, sous une forme électronique » (Bonnaud-Le Roux, 2020). Appliquée au patrimoine, la notion de numérisation renvoie à une palette de techniques permettant de dupliquer un bien culturel pour aboutir à sa représentation numérique

10. Voir The Future of Life Institute (FLI), « Asilomar AI Principles », 2017, <<https://futureoflife.org/2017/08/11/ai-principles/>>, (consulté en février 2022).

réalisée à l'aide d'un logiciel informatique de modélisation. Elle se décline, aujourd'hui, au sein des différentes composantes du patrimoine.

1.2 Usages de l'intelligence artificielle par les guides qualifiés

Sites Internet, applications, pages Facebook, comptes Instagram et Twitter, balados (*podcasts*) et vidéos en ligne, écrans tactiles, audioguides de nouvelle génération, expériences de réalité augmentée, expositions immersives et visites virtuelles, numérisation des catalogues et des archives, etc., les outils numériques auxquels les musées peuvent faire appel sont légion. De la communication à la conservation, de la médiation à la recherche et aux acquisitions, tous les secteurs du musée sont concernés (Guidi, Jenny et Hagen, 2021) : communication, médiation, publics, archivage, collecte, préservation et scénographie.

Les expériences en lien avec le numérique abondent : l'oculométrie, système de suivi du regard permettant de concevoir un audioguide personnalisé en fonction de ce qui est scruté dans l'œuvre ; la numérisation en très haute définition afin d'animer les objets ; l'analyse des collections en se basant sur les couleurs que le visiteur ou la visiteuse choisit dans un cercle chromatique ; la classification d'images en utilisant un outil de reconnaissance de formes dans les images mêmes et permettant ensuite de leur attribuer des mots clés ; la numérisation de la bibliothèque permettant de faire des recherches par textes et, à long terme, la volonté d'intégrer également la reconnaissance par images afin d'identifier, d'une manière plus précise, les profils des visiteurs et des visiteuses pour mieux cerner leurs besoins.

Ainsi, au moyen de ces outils, il est proposé aux visiteurs et visiteuses de vivre des expériences virtuelles, en ayant une meilleure compréhension des contenus combinant les images animées et des textes. Chaque personne qui accepte de s'approprier l'espace proposé et ses objets y laisse une trace de sa personnalité, qui permet en retour au concepteur de contenu de proposer des services adaptés. Il existe également plusieurs moyens pour retracer les comportements des usagers : statistiques, études des publics et de leurs mouvements, durée pendant laquelle ils regardent l'exposition et ce qu'ils regardent (d'où une adaptation du parcours en fonction des intérêts) si les publics se connectent aux QR codes proposés, par exemple.

1.3 Comment l'IA peut s'inscrire dans les espaces de la visite guidée : ancrages et conditions

L'espace de la visite guidée s'ouvre à partir d'une *relation* qui tient ensemble le visiteur, *le visité* et leur interprète, relation matérialisée dans un espace-lieu¹¹. Telle configuration désigne une forme de communication humaine qui implique fondamentalement une transmission d'information spécifique. Cette information devrait être perçue en tant que telle par les récepteurs et réceptrices, c'est-à-dire qu'elle devrait modifier leur comportement par une révision de leurs croyances relatives au monde. C'est par cette confrontation que se considèrent également les caractéristiques de la médiation culturelle, laquelle mobilise des moyens pour aborder le patrimoine par un processus d'appropriation et de son redéploiement (Chaumier et Mairesse, 2013, p. 52). Enfin, nous savons que l'espace de la visite est un espace délimité : temporellement, matériellement et cognitivement (Cohen-Azria et Lahanier-Reuter, 2013, p. 61). Par ce dernier élément, la langue comprise comme une dimension de la culture et de la civilisation acquiert du poids. Ce caractère à la fois transnational, lucratif et ponctuel de l'activité est également reconnu par la jurisprudence de l'activité de guide. Désormais une visite guidée, assurée par le même guide, peut être délivrée partout en Europe si elle ne menace pas la sécurité du territoire des États membres.

Les guides se situent alors au centre des processus décrits et la manière dont ils participent à l'échange des biens et services nécessite une qualification et ne peut exister sans un dialogue soutenu entre le monde de l'enseignement, du travail et de la législation. À partir de ce constat, et considérant l'art du guide reconnu en tant qu'action professionnelle établie dans le champ de la médiation patrimoniale, à travers l'interprétation et la représentation (dans un musée ou sur un site protégé), il nous est possible d'observer comment la politique publique de l'Union européenne est transposée dans ce type d'action.

De là, nous déduisons que le guidage passe par la maîtrise d'un langage approprié et adapté à la situation rhétorique (Bitzer, 1968), pour garantir que les actions entreprises des visiteurs et visiteuses, lors de la visite guidée, respectent les règles du fonctionnement local. Les modalités de

11. Ce sont des espaces publics/privés, intérieurs/extérieurs, des lieux de savoir, des lieux de pouvoir étatique (Parlements, sièges, bureaux), des lieux de mémoire, l'espace public (rues, quartiers, villes, régions, pays), les transports (bateaux-mouches, croisières, bus, voiture...).

l'activité sont façonnées par les attentes des visiteurs, mais également des visités. Le guide assure la transmission des valeurs aux publics spécifiques de la visite guidée (Gellereau, 2006). De cette manière, les services du guide revêtent un aspect didactique, avec une notion prononcée d'aide à la personne. Le guide est individuellement responsable de la relation interpersonnelle entre visiteur et visité. Son discours est doublement orienté par les attentes des individus et par la situation rhétorique (Lebel, 2003). Cette dernière dépend fortement de son statut d'emploi et de la représentation que nourrit le guide de sa compétence. De plus, il s'agit de messages codés, reçus, interprétés et retransmis dans le processus plus vaste du patrimoine. Cette expérience s'appréhende en majorité par les pratiques langagières. Discours et langue de travail du guide relient les expériences patrimoniales en vue d'une appropriation du lien à la mémoire que l'on cherche à conserver et à protéger. Autrement dit, c'est en se déplaçant dans l'espace (par les commentaires autour des objets et des processus) et en observant le paysage qu'on lui donne du sens et, par une diffusion contrôlée de l'information, qu'on le préserve et le valorise. Cette observation passe par des images verbales, dotées d'une puissance symbolique que suscite ainsi la situation rhétorique (Charland, 2003).

L'observation de l'activité du guide dans son milieu naturel démontre sa forte dépendance à l'égard du milieu institutionnel, autant par son statut d'emploi que par le contenu et l'organisation des visites guidées. À l'occasion de la visite guidée se vit une expérience patrimoniale. L'expérience patrimoniale se caractérise par le lien à la culture locale. La visite guidée, qui la situe dans le temps et sur un territoire, peut faire l'objet d'une normalisation (Turinetti, 2018, p. 18).

Définir la compétence du guide dans ce contexte sert à comprendre la logique du processus de normalisation – lequel a pour objet la formalisation des savoirs, des règles de l'art et des usages techniques. Ces règles, en vue de normaliser les pratiques et d'harmoniser les comportements, sont fixées par les corporations de métiers (Balibar, 2013)¹². Ainsi, une profession organisée autour de l'expérience patrimoniale de l'accompagnement, qui représente une échelle modeste de professionnels indépendants,

12. Étienne Balibar affirme que la dérégulation généralisée est une expression qui désigne une vision strictement économique du monde de produits standardisés, où intervient la volonté politique spécifique, inquiète par la construction du mécanisme de gestion et de régulation. Cette idéologie est sujette aux contradictions internes, qui tendent à maximiser les forces centrifuges au lieu de créer une volonté commune.

devient un lieu d'observation privilégié des innovations administratives et numériques. Il est intéressant d'observer, à ce stade, la manière de définir la singularité: l'individu singulier, pour le législateur européen, peut être le consommateur dans le monde de la concurrence. Ici, la singularité désigne un choix de vie. L'orientation envers le consommateur provoque l'adaptation du contenu, qui laisse dans l'ombre le geste de la régulation. L'individu devient ce qu'il consomme; s'il ne consomme plus, il disparaît. De cette façon, nous pouvons observer les nouvelles manières d'agir, de penser, de regarder le monde. De fait, la pratique de guide persiste dans ces espaces, mais elle manque d'une niche dans la communication médiatisée et risque d'être remplacée par la mise en place d'un récit épuré.

2. L'EFFET PRODUIT PAR L'INTRODUCTION DE L'INTELLIGENCE ARTIFICIELLE SUR LA DÉONTOLOGIE DE LA PROFESSION

Le fait culturel est d'abord un fait d'institution. Les institutions sont des systèmes de normes, de règles, de hiérarchies de langue et de langage, des systèmes de croyance collective. Aujourd'hui, les usagers au sein de ces systèmes sont reliés par les plateformes et les applications du Web, qui représentent en quelque sorte son organisation sociale (Hui, 2017). Dans le cas des guides, il peut parfois s'agir d'une organisation interne, régie par des normes internationales exécutées par les plateformes collaboratives:

Dans le monde entier, nous travaillons soit avec les ministères du tourisme, soit avec des syndicats, soit avec des associations de guides... 2500 guides, dans 174 pays, font partie de [notre plateforme] aujourd'hui [2016]. Nous comptons monter jusqu'à 20 000 guides inscrits sur les 200 000 qui exercent dans le monde. Nous n'irons pas au-delà pour maintenir des standards de qualité. Vous me dites: je veux un guide au Tadjikistan qui est disponible samedi prochain de 17 h à 18 h et qui parle français, aujourd'hui je ne suis pas sûr que vous puissiez tout de suite appuyer sur un bouton et avoir un guide à ce moment-là. Trouver un guide qui est disponible samedi prochain et qui parle slovène à Paris, à travers [les plateformes numériques], c'est devenu possible. Jusqu'à ce moment, personne n'a encore réussi à commercialiser ni à vendre sur Internet, ni sur le mobile, les prestations de guidage. Mais nous avons identifié le besoin d'un appui, nécessaire au guide, notamment aux niveaux commercial et administratif, au niveau logistique, au niveau du service client et surtout pour toucher un nouveau cycle de clientèles qui vient des nouveaux supports. Ces nouveaux supports sont le site Internet et le mobile (extrait d'une observation, Uzlyte, 2021).

Le modèle de travail décrit par la citation plus haut n'est pas encore dans les usages, mais il persiste comme une pratique numérique permettant de collecter d'abord les données (noms, prénoms, images, langues parlées) afin de favoriser la flexibilité dans le projet de déplacement. Désormais, il est possible d'engager un guide quasi instantanément en lui demandant telle prise en charge. Et les plateformes collaboratives¹³ font évoluer les formes organisationnelles de la profession en visant à formaliser et à normaliser des situations professionnelles *in situ*.

Dans cette perspective, « l'informatisation est la forme contemporaine de l'industrialisation » (Volle, 2012, p. 87) qui restructure les pratiques sociales de telle manière que l'informatique devient inséparable de la gestion et de la production (Vial, 2013). De là émerge un nouveau milieu de communication, soit un système de diagnostic automatisé (Sadin, 2005, p. 91, 142), composé d'algorithmes *ad hoc* agrégés aux dossiers individualisés et exploités dans l'objectif de monnayer des biens et services, pour suggérer ensuite des « feuilles de route » personnalisées. Cela se fait par un projet d'automatisation (Hui, 2017, p. 169), qui consiste en trois étapes : 1) la création des standards ; 2) leur application aux « données brutes » pour acquérir les sémantiques ; 3) l'implémentation des « agents », c'est-à-dire des algorithmes qui analysent leur sens.

Yann Le Cun (2019, p. 251) précise les types d'IA en application et en distingue quatre. Le premier, l'apprentissage profond (*deep learning*) ou « apprentissage supervisé », permet de percevoir l'environnement et d'autres modules de planification de trajectoires qui tiennent compte de cartes détaillées, établies à l'avance. Ce type d'IA, soit l'analyse à grande échelle des interactions sociales, repose sur le principe d'une architecture dont les paramètres s'adaptent progressivement pour approcher la tâche demandée. Cet apprentissage supervisé est très efficace lorsqu'on dispose de suffisamment de données ; mais, à l'intérieur d'un périmètre donné, il conserve des zones aveugles, notamment en lien avec le problème de la quantité des données recueillies.

Le deuxième poursuit Le Cun, l'« apprentissage par renforcement », permet d'entraîner une machine sans lui donner la réponse attendue, en lui indiquant seulement si la réponse produite est correcte ou non. Mais

13. La plateforme numérique dont il s'agit (Meetrip) est un lieu de vente virtuel des services des guides. Elle peut proposer 10 000 guides professionnels dans plus de 2000 villes et 230 pays. Elle a une fonction de coordination et de mise en réseau.

l'IA n'a qu'une compréhension limitée aux exemples appris du monde. Un système de traduction fait, parfois, de grands contresens sans en avoir la moindre idée. Ce système est centré sur la perception et l'interprétation de signaux qui visent l'architecture générale d'un agent intelligent, autonome, capable d'apprendre à percevoir et à planifier. Ici, deux mécanismes entrent en jeu. Le premier, réactif, de type stimulus-réponse, commande les tâches que nous pouvons exécuter sans réfléchir. Le second fait intervenir notre modèle du monde. Ces deux modes de pensée et d'action sont catégorisés par « système 1 » et « système 2 ». Les comportements de type 1 sont soit innés, soit appris. Les comportements de type 2 sont raisonnés, conscients et réfléchis. Dans le premier cas, la pensée est rapide, instinctive et émotionnelle. Le deuxième cas concerne, à l'opposé, la pensée lente et délibérative. La nature de la communication numérique permet rarement ce deuxième mode de pensée, car l'interaction virtuelle est habituellement rapide, éphémère et émotionnellement intense. Les systèmes de l'IA relèvent plutôt du système 1. Il va de même pour le troisième type d'IA qui pourrait être distingué comme « intelligence artificielle universelle ». Ici, les systèmes choisissent une action qui maximise la récompense attendue, compte tenu de l'information qu'ils ont acquise. Mais cela reste encore difficilement réalisable en pratique. Enfin, le quatrième type, spécifié par Shanahan Murray (2018, p. 93), pourrait porter le nom de « superintelligence artificielle » ou encore IA abrégée (AGI) qui vise à égaler les humains dans (presque) toutes les sphères de l'activité intellectuelle ; il suffirait qu'elle utilise une combinaison de processus d'optimisation et d'algorithmes d'apprentissage automatique incorporant une compréhension sensée du monde et permettant l'émergence de la créativité. Contrairement au cerveau biologique, l'émulation cérébrale réalisée numériquement peut être copiée (et recopiée) en quantité arbitraire. Et, contrairement au cerveau biologique, le cerveau numérique peut être accéléré. Ce qui est commun aux quatre cas observés, c'est la récolte et la conservation massives de données se développant par défaut : les mégadonnées (*big data*).

Ces données proviennent de partout : les détecteurs (sondes) utilisés pour la collecte d'informations météorologiques, les publications sur les réseaux sociaux, les photos et vidéos numériques, les bulletins d'ordres de transaction, ainsi que les signaux GPS des téléphones portables. [...] [A]insi étant face à la situation de rareté de l'information et de difficulté à trouver des données, nous sommes désormais confrontés à un « déluge de données » (Sorin Calude et Longo, 2017).

Les auteurs cités soulignent que ces données sont traitées au moyen de corrélations, sans analyse causale ou sémantique. Par exemple : au sein de l'Institut de recherche et d'innovation (IRI) sont largement expérimentées les technologies contributives reposant sur des processus d'annotations contributives, elles-mêmes fondées sur des processus de métacatégorisation. Pour nous rendre compte comment cela a des répercussions sur l'activité de guide, arrêtons-nous un instant à la définition de l'IA proposée par la Commission européenne. Elle décrit l'intelligence artificielle en tant que logiciel ou programme « qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit¹⁴ ». Les données rassemblées nous permettent d'observer une émission d'une certaine réserve envers la régulation ou l'encadrement par des lois considérées comme trop strictes. Selon les propos recensés, les guides misent davantage sur des règles de bonne conduite et la production de normes techniques renouvelées tous les dix ans.

2.1 L'évolution du code de conduite de la profession de guide

L'évaluation des pratiques professionnelles est traitée à l'aide des algorithmes, même si ces décisions ont des dimensions et des conséquences d'ordre moral. En apparence, les algorithmes semblent objectifs, mais nous n'ignorons pas qu'ils sont façonnés par des ingénieurs pouvant avoir leurs préjugés (Barnet, 2019, p. 35). Nous savons que la détermination d'intention est indispensable pour traiter la recherche d'information. Une IA, programmée pour accumuler le plus de ressources possible, transforme tout en matière programmable (Murray, 2018, p. 175). Il s'agit là d'un « but instrumental convergent ». « Convergent », parce qu'il est susceptible de se retrouver dans toute intelligence artificielle générale avancée, dont la fonction de récompense est ouverte et non triviale. « Instrumental », parce qu'il n'est qu'un moyen de parvenir à une fin, et non pas une fin

14. La proposition de règlement faite en avril 2021 s'appuie sur l'article 114 du Traité sur le fonctionnement de l'Union européenne (TFUE) concernant les mesures destinées à renforcer le marché intérieur européen. La Commission fonde également ces dispositions en droit sur l'article 16 du TFUE qui dispose que « toute personne a droit à la protection des données à caractère personnel la concernant ». Voir CE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union projet de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle*, dit « *Artificial Intelligence Act* », COM (2021) 206 final, 21 avril 2021.

en soi. La fin en soi du système est de maximiser une certaine fonction de récompense. Un autre but instrumental convergent est l'acquisition de ressources. Argumentations et contre-argumentations tiennent pour acquis le lien entre l'intelligence et les droits. Pour jouir de ces droits, l'IA devrait d'abord être capable d'agir sur le monde (Vial, 2013, p. 32) :

Si la prise de conscience a toutefois commencé, nous le devons moins aux philosophes qu'aux institutions, dont l'ouverture à la culture technique remonte au début des années 1970 avec la création, au sein du Centre Georges Pompidou, à Paris, du Centre de création industrielle. [...] Dans les décennies suivantes, les produits de la technique s'intègrent à la programmation des musées, jusqu'à faire naître des institutions dédiées comme la Cité des sciences et de l'industrie, créée en 1986, ou faire entrer des produits high-tech dans les plus hauts lieux culturels, comme lors de l'exposition *Game Story* : une histoire du jeu vidéo, programmée au Grand-Palais à la fin de l'année 2011.

Ce qui est décrit plus haut est défini par Stéphane Vial comme une matrice ontophanique, c'est-à-dire une structure générale de la perception qui conditionne *a priori* que les êtres apparaissent en tant que données. Reconnaître ce fait, c'est admettre que nous sommes face à un « système de domination qui fonctionne au niveau même des conceptions et des constructions des techniques ». En d'autres termes, l'appareil s'impose à la pensée et à la manière d'être au monde ; il devient une conduite d'existence ; nous n'existons que si nous faisons usage de tel appareil ou nous exprimons avec telle technique. Vial poursuit que le comportement décrit correspond à ce qu'engendre le phénomène numérique : la technique désigne une structure de perception et conditionne la manière dont le réel ou l'être nous apparaît. Or, la perception n'est pas uniquement une fonction physique ou cognitive, « elle est aussi une fonction sociale [...] conditionnée par des facteurs culturels » (Vial, 2013, p. 135). Par exemple, un virtuel psychique, qui « [...] concerne l'ensemble de nos attentes et de nos représentations préexistantes à une rencontre réelle ». Ce virtuel psychique est « un monde interne de représentations fantasmatiques que nous attachons à un être réel et qui, lorsqu'une rencontre avec cet être réel se produit dans la réalité concrète, doit être physiquement retravaillé afin de tenir compte de cette réalité » (Tisseron cité par Vial, 2013). Autrement dit : le virtuel est un ensemble d'images informatiquement simulées, qui forment la base des langages de programmation permettant de réduire un problème à un algorithme.

À cela Vial ajoute que les appareils numériques ne pouvant être utilisés qu'à l'aide d'interfaces, la nature de l'expérience opérationnelle qu'ils offrent à l'utilisateur n'est pas de l'ordre de l'action mécanique, mais de l'interaction algorithmique. Lorsqu'on utilise un ordinateur, on n'agit pas : on interagit (Vial, 2013, p. 206). Vivre dans ces interactions, c'est donc vivre une relation potentiellement infinie avec la matière calculée. Seul ce qui est fait de matière calculée peut engendrer de l'interactivité, c'est-à-dire de l'activité corrélative ou, dit autrement, stimuler le décalage entre le déclaratif et la réalité.

Si les liaisons sociales activables au sein d'un groupe dépendent des appareils, selon le système technique dans lequel nous vivons, nous n'élaborons pas la même relation à autrui, parce que, poursuit l'auteur, nous utilisons des appareils différents. Nous en déduisons qu'on ne peut pas nier que les programmes nous aident à communiquer, mais il s'avère que nous prêtons aux machines plus d'intelligence qu'elles en ont réellement. Pourquoi ? D'abord parce que, pour l'homme, le langage est si fondamental qu'il est synonyme d'intelligence (Le Cun, 2019, p. 366). Nous associons des mots à des concepts, nous les manipulons pour produire des raisonnements. Tout cela nous amène à penser que l'intelligence ne peut exister sans langage codé. De plus, l'humanité a une longue expérience de la codification de ces systèmes de valeurs morales, dans les lois d'une part ; ces valeurs sont également codées par l'éducation, d'autre part. L'intelligence ne se réduit donc pas aux seules facultés intellectuelles : elle touche à toutes les sphères du comportement ; elle est à la fois apprentissage, adaptation et faculté de prise de décision.

C'est alors qu'il serait important de nous défaire des tendances anthropomorphiques dans la réflexion menée sur l'intelligence artificielle générale, car nous risquons d'oublier qu'une IA utilise un langage à des fins purement instrumentales. Nous avons vu qu'en tant qu'individus nous sommes dépendants des technologies de l'information – et l'intelligence artificielle sophistiquée ne fera probablement qu'accroître cette vulnérabilité (Jorion, 2017, p. 63-76). En ce sens l'aspect social du guidage démontre les effets de cette restructuration sur une profession culturelle à travers les processus de rationalisation.

Une éthique bidimensionnelle émerge dans ce contexte de la visite guidée comme un point de jonction entre système d'information et système d'administration publique du pouvoir. Elle apparaît comme une éthique relationnelle – celle qui encadre et favorise les modes d'accord

dans les organisations, à des fins d'analyse et d'aide à la décision. Elle joue alors un rôle de croyance, où les savoir-faire sont reformulés en faire-savoir. Puis comme une éthique professionnelle appliquée à l'action du guide, qui devient une ambition : celle de maintenir un espace de dialogue qui se présente sous forme de discours et de pratiques normatives en confrontation. Ces discours, précisons-le, véhiculent une idéalisation de la transparence et de la participation égalitaire aux informations, donc aux connaissances. Les sources d'information constituent un élément déterminant dans la réception de l'information. La visite guidée permet de maintenir cet espace public médiatique ouvert.

Le code de conduite professionnelle est une des réponses à cette interrogation de l'ordre éthique. En effet, produire une charte suscite des interactions entre les polarités opposées. C'est une source de tensions, mais aussi de socialisation et de reconnaissance mutuelle. Ainsi les codes de conduite représentent un nouveau système de régulation sociale moins restrictif que la réglementation, basé sur la flexibilité des engagements. Ces codes sont établis en fonction des exigences des interlocuteurs et des régulations souhaitées.

2.2 Évolution de la pratique de la visite guidée avec l'introduction de l'intelligence artificielle dans le domaine du patrimoine

La visite guidée au musée apparaît comme une expérience, souvent dans le sens du *naturalistic humanism* de John Dewey (Dewey, 1958), qui lie l'expérience à deux notions, celle de situation et celle d'interaction qui a lieu quand la transaction entre l'individu et son environnement se produit. Ici l'expérience façonne un *continuum* (Montpetit, 2005). Cette cohérence permet d'inscrire la visite guidée en tant que dispositif d'éducation formelle ou non formelle, comme un moyen optimisé favorisant la cohésion sociale. La visite guidée se présente comme l'occasion d'une relation interpersonnelle, comme une expérience didactique et déontologique qui encourage les participants à accepter « de nouvelles valeurs » et à développer leur sentiment d'appartenance à la communauté. Regardée dans ce sens, l'activité de guide intervenant au musée (pour les groupes d'élèves) se définit de la manière suivante :

L'essentiel du travail de guide est de présenter et de démontrer la valeur et la signification historique, scientifique, idéologique et esthétique de l'héritage culturel, à l'aide d'objets, de groupes d'objets, d'images et de documents. Dans les faits, la visite est une activité extrascolaire qui se déroule

à l'occasion de fêtes ou à l'issue d'un concours. Malgré tout, la visite doit respecter le contenu d'une leçon en classe. Au cours de la visite, les élèves écoutent attentivement le guide, admirent les objets, prennent des notes et posent des questions (Allard et Lefebvre, 2001).

Nous constatons trois contextes : personnel, social et physique (Falk et Derking 1994, p. 133 ; 2000), cadre qui est aussi celui de l'expérience culturelle. L'expérience culturelle décrite peut s'appréhender comme un comportement spécifique : les visiteurs consomment l'expérience qui fait partie du produit culturel industrialisé ou en est parfois la base (Holbrook et Hirschman, 1982). Et le comportement du consommateur de culture y est largement étudié, mesuré, estimé (Mairesse et Rochelandet, 2015, p. 180). Dans ce contexte, les guides se situent parmi le personnel de contact. Ainsi la visite devient un cadre d'échanges entre l'établissement, le territoire et le visiteur (dans un contexte touristique) :

Dans toute visite, il existe un système d'échange où trois éléments sont à prendre en considération : d'abord, la participation du bénéficiaire (le touriste), ensuite l'impact du service rendu qui met en cause le fonctionnement économique du musée ou du monument historique dans une stratégie de développement territorial, enfin la qualité de la prestation qui sera évaluée par un consommateur final (le visiteur) (Mairesse et Rochelandet, 2015, p. 182).

Il s'agit dans ce qui précède de l'« orchestration » des dispositifs législatif, réglementaire et conventionnel. Ainsi, le guide apparaît comme un outil de l'interprétation, où la visite guidée est la plus-value d'un produit touristique culturel¹⁵. Il cherche à transformer l'espace phénoménal en « représentation de l'espace », à l'aide d'un accompagnement professionnel dont l'essence est l'individualisation. À partir de là, nous voyons se dessiner une architecture du métier, expression de son économie :

1. Pratiques exemplifiées,
2. Institution de la visite, traduite par le rapport entre le statut de la visite guidée et celui du guide,
3. Cadres socio-politico-économiques où s'inscrit la profession,

15. Dans la *Convention de Faro sur la valeur du patrimoine culturel pour la société* adoptée par le Conseil de l'Europe en 2005, nous lisons : « Un ensemble de ressources héritées du passé que des personnes considèrent, par-delà le régime de propriété des biens, comme un reflet et une expression de leurs valeurs, croyances, savoirs et traditions en continuelle évolution. Cela inclut tous les aspects de l'environnement résultant de l'interaction dans le temps entre les personnes et les lieux », art. 2. a).

4. Procédés, fonctions et représentations symboliques des guides,
5. Bonnes pratiques et échanges de l'expérience qui enseignent les comportements et les compétences.

L'observation des réunions interministérielles de concertation, autour d'une nouvelle approche de la profession et du patrimoine, que nous avons effectuée en 2015-2017 en France, nous a permis de constater que, pour cet exercice d'évaluation en vue de la normalisation proposée par la Commission européenne, trois critères sont à considérer : la non-discrimination, la nécessité et la proportionnalité au regard des objectifs d'intérêt général. Ainsi la profession de guide a été passée en revue pour observer si toutes ses conditions d'accès et d'exercice étaient justifiées, nécessaires et proportionnées au but d'intérêt général poursuivi. Dans cette évaluation, deux interrogations problématiques ont émergé : 1) En quoi les tâches réservées à la profession de guide sont-elles assez complexes pour justifier la possession d'une qualification professionnelle déterminée ? ; 2) Les qualifications professionnelles sont-elles adaptées à la complexité de ces tâches ? Ces deux interrogations forment l'épicentre du débat européen autour de l'organisation de la profession de guide.

La normalisation est généralement définie comme « un processus d'élaboration et de production de documents de références, c'est-à-dire de normes¹⁶ » (Grenard, 1996, p. 45). On désigne par norme – un commandement, un ordre, une prescription. Commander n'est pas l'unique fonction d'une norme ; c'est aussi habilitier, permettre, abroger. Dans la mesure où la « norme » désigne une prescription ou un ordre, cela signifie que quelque chose doit être ou doit avoir lieu. Le devoir-être est une volonté dirigée vers le comportement d'autrui, d'un acte dont la signification est qu'une autre personne ou un ensemble de personnes doit se comporter d'une manière déterminée qui mène vers la question de sa validité (Kelsen, 1996, p. 56).

Dans la définition de l'objectif d'intérêt général retenu par la Cour de justice de la Communauté européenne, il ne s'agit pas de la protection du patrimoine, mais de la meilleure diffusion possible des connaissances relatives au patrimoine. Ainsi, la profession structurée par la professionnalisation apparaît comme dépendante des instances politiques, non

16. Agnès Grenard précise que la fonction spécifique d'une norme est le commandement d'un certain comportement : « commander », « prescrire », à la différence de « décrire ». Décrire est la signification d'un acte de connaissance, commander ou prescrire, d'un acte de volonté ; on décrit quelque chose, en énonçant ce qu'il est ; on prescrit quelque chose, en exprimant ce qu'il doit être.

seulement par son statut professionnel, mais également par son activité, qui est l'interprétation du patrimoine et la transmission des savoirs.

Par les réformes entamées autour des guides européens qualifiés, la profession est amenée à s'organiser, par la mise en place des programmes de formation professionnelle et de qualification des guides touristiques. Nous le comprenons comme une attention prêtée à la reconnaissance de la profession qui, dans le vocabulaire juridico-administratif, est la pratique qui consiste à influencer les politiques et le processus décisionnel, en vue d'obtenir un résultat précis. Les guides accomplissent cette action en négociant les normes avec le législateur national et le législateur européen.

Depuis, plusieurs vagues de réformes autour du métier de guide ont suivi en Europe : la première sur le statut autour de 1995-2000, la deuxième sur la définition du champ d'action en 2009-2010 et la troisième sur la définition de la compétence et du système de formation, à partir de 2015. Dans le troisième cas, il s'agit d'un portfolio pour évaluer l'acquisition des compétences normalisées. Deux dates sont ici à retenir¹⁷ :

En 2003, par le biais de la FEG et au sein du Comité européen de normalisation (CEN), une définition a été élaborée. Cette définition était cruciale pour nous, parce qu'en Europe elle a été adoptée par différentes institutions [et maintenant] c'est une norme ISO (une norme internationale). En 2008 (CEN 15565), il y a eu une deuxième norme concernant la formation des guides touristiques. Ces deux normes européennes sont uniques actuellement, elles sont adoptées par l'association européenne (European Federation of Tourist Guide Associations, FEG), et l'association mondiale (World Federation of Tourist Guide Associations, WFTGA) et acceptées par toute la communauté professionnelle des guides qualifiés à l'échelle mondiale (Entretien, 2017, Uzlyte, 2021).

Imaginons maintenant, pour illustrer ce qui est présenté plus haut, un guide non qualifié qui a suivi une formation à distance sur un site offrant des outils de communication adaptés à un guidage efficace. Il est devenu un usager expérimenté de ce site et ses visites, interactives, séduisantes et ludiques, ont beaucoup de succès. Mais il n'a pas acquis de connaissances approfondies du contexte politique, économique, social ou symbolique

17. Du point de vue de la Fédération mondiale des associations de guides touristiques et d'autres, comme la Fédération européenne, la Fédération d'Asie du Sud-Est, l'Association des pays arabophones, durant ces trente dernières années, deux événements ont marqué la vie professionnelle des guides : deux normes européennes (2003, 2008) qui visent à définir l'activité de guide et à la structurer.

où s'inscrivent ces visites, ni des lieux qu'il représente, ni des particularités du public qu'il est en train de guider. Ainsi, c'est un guide fiable pour l'administrateur du site, mais nous ne pouvons pas le qualifier de guide professionnel, en raison de son manque de compréhension plus profonde de ses actions. Savoir, savoir-faire, savoir-être s'acquièrent, aujourd'hui, par des processus d'apprentissage à la fois formels et non formels tout au long de la vie, encadrés par la Commission européenne et le CEDEFOP¹⁸. Par exemple, la proposition de normaliser la grille de formation est désormais connue en tant que norme EN 15565¹⁹, à propos des exigences relatives aux programmes de formation professionnelle et de qualification des guides touristiques. Il s'agit d'un programme ciblé et adapté autant pour habiliter les enseignants que pour accréditer les étudiants²⁰, comme dans le cadre du projet Innoguide²¹ :

Le projet vise à introduire une innovation dans le programme de formation des guides touristiques. Selon une définition standard, un guide touristique est un professionnel ayant le potentiel de contribuer au développement d'une région. [...] Les guides constituent un groupe professionnel important dont la formation a besoin d'innovation pour pouvoir contribuer au changement menant à une Europe plus forte, ou à la dimension européenne (DEE). Les guides peuvent contribuer au changement grâce à l'éducation informelle qu'ils offrent aux touristes, une manière innovante qui peut contribuer aux objectifs du programme LLL – la connaissance – et ainsi influencer le changement des comportements et des attitudes²².

-
18. Centre européen pour le développement de la formation professionnelle (CEDEFOP) dont la mission consiste à soutenir la Commission européenne dans le développement des politiques de formation professionnelle et technique en Europe et de contribuer à leur implantation.
 19. Voir la norme NF EN 15565 : « Services touristiques. Exigences relatives aux programmes de formation professionnelle et de qualification des guides touristiques », qui spécifie les exigences minimales relatives aux programmes de formation professionnelle et de qualification des guides touristiques. En l'absence de commission de normalisation compétente, et après avoir sollicité l'avis des acteurs socioéconomiques concernés, le présent document a été homologué comme norme française et sa référence a fait l'objet d'une publication au JO. Voir *Avis relatif à l'homologation et à l'annulation de normes*, JO, 8 mai 2008, n° 0108.
 20. Les formations de la FEG intégrant la norme sont recensées, accessibles en ligne. European Federation of Tourist Guide Association, « Seminars & Courses », <https://www.feg-touristguides.com/training_seminars.php>, (consulté en août 2021).
 21. The INNOGUIDE is a Leonardo da Vinci Life Long Learning (LLL) Program sponsored by the European Union (EU) under the project « Innovation in LLL Tourist Guide Training, Tourist Guides as Partners for a Sustainable, Diverse, and Exciting Europe ». Project n° 510487-LLP-1-2010-BE-LEONARDO-LMP.
 22. La recherche INNOGUIDE analyse d'abord les niveaux du Cadre européen des certifications (CEC), ainsi que la structure et le contenu des programmes d'études des établissements de formation nationaux ou régionaux des huit pays européens participants : Belgique, France,

Les pratiques de régulation orientées vers les droits de consommateur et la concurrence, portées par la Commission européenne, confirment le contexte socioéconomique néolibéral. Dans ce contexte, la communication numérique « lisse » la notoriété des établissements, donc les comportements des professionnels qu'ils emploient (Corral-Regourd et Alcantara, 2016). De ce fait, la compétence de guide qualifié intervenant dans ces établissements évolue. En revanche, reconnaissant que toute visite revêt une dimension culturelle, donc sociale, il devient possible de situer l'activité du guide qualifié dans le domaine de la médiation culturelle et développée comme une pratique d'accompagnement des visiteurs et des visiteuses. Compte tenu des compétences spécifiques requises pour exercer l'activité et du lien étroit avec le marché, de nouvelles formes d'apprentissage tendent à orienter les pratiques et visent des modèles de comportements formalisés.

Dans les musées ou les monuments historiques, nous pouvons observer une stratégie de médiation nouvelle – apporter un simulacre de déplacement et des effets visuels (Renaud, 2022)²³. Ainsi, un glissement de la prétention communicationnelle, du syntagmatique au paradigmatique, a pu être constaté : « Dans le sillage des travaux de Charles Sanders Peirce, de la *semiosis sociale* d'Eliseo Veron et des *social semiotics* britanniques, nous considérons que la production du sens ne peut être analysée indépendamment du sujet qui perçoit et interprète » (Saemmer-Cemti et Tréhondart, 2022). De l'exemple du Centre des monuments nationaux (CMN) en France, nous pouvons cerner les objectifs d'une telle médiation renversée provenant des établissements culturels publics : « identifier comment satisfaire les attentes des établissements scolaires, identifier quelles pratiques diffuser, positionner l'offre du CMN, identifier les partenariats pour la meilleure distribution, technique et médiation » (Bergeot, 2022). Cependant, il apparaît difficile de savoir comment cette évolution, dans les années à venir, influencera

Hongrie, Italie, Malte, Norvège, Espagne et Pays-Bas. Ensuite, INNOGUIDE construit une analyse comparative pour examiner dans quelle mesure ces aspects sont prédominants dans les programmes de formation des guides actuels. Afin d'introduire des innovations dans le guidage et d'atteindre les objectifs d'éducation et de formation tout au long de la vie, les résultats seront utilisés pour développer trois modules de formation de guides touristiques : durabilité, interculturelisme et guidage expérientiel.

23. Lise Renaud met en exergue une logique documentaire (inventaire, classification et identification), une logique panoramique (modélisation), une logique spectaculaire (effets visuels), une logique du stimulation (reproduction d'une pratique sociale) et propose une typologie des archétypes d'expositions en ligne : écran-catalogue, écran-cartel, écran-maquette, écran-médiateur, écran-diorama, écran-scénographie.

la manière de partager la culture européenne, et peut-être aussi une certaine idée de la médiation culturelle.

CONCLUSION

Il ne faudrait pas oublier que les algorithmes sont des instruments installant un environnement utilitariste et réclamant, en retour, de leur utilisateur un rapport instrumental (Cardon, 2017, p. 38). Et ce rapport est tout autre que celui qu'on rencontre lors de la visite guidée plus communément répandue, assurée *in situ* par un guide qualifié. Les deux pratiques nécessitent d'être comparées pour comprendre que les deux pratiques ne relèvent pas des mêmes catégories de l'activité culturelle ; elles sont complémentaires, certes, mais difficilement comparables comme des expériences équivalentes. Néanmoins, à partir de cette confrontation et des données recueillies concernant la pratique des guides experts, nous pouvons déduire deux enjeux principaux pour la profession. 1) Les innovations administratives ont permis de structurer l'action de guide que nous pouvons aujourd'hui considérer comme une profession, ce qui n'était pas le cas il y a soixante ans (les premiers syndicats sont apparus, les formations se sont mises en place, le cadre des interventions a permis le travail salarié ; enfin, le statut professionnel et social de guide a dû être arrêté). 2) Les pratiques culturelles numériques sensibilisent le public et, au lieu de « prendre » les parts de marché, elles l'élargissent et encouragent les visiteurs à se déplacer physiquement : « Parmi les 9 % des Français déclarant avoir consulté le site d'une exposition ou d'un établissement patrimonial en 2018, environ la moitié déclare avoir aussi visité un site patrimonial *in situ*. A contrario, 4 % des citoyens se connectent uniquement en ligne à la culture institutionnelle » (Appiotti, Ballarini et Bernon, 2022).

Ici, il est souhaitable de ne pas oublier non plus que, derrière chaque icône et chaque bouton, existent des lignes d'algorithme (Vial, 2013, p. 196) et que nous sommes face à des systèmes de prescription automatisés où il devient difficile de discerner quel intérêt s'exprime. La fonction du système informatique se limite à la mise en place de l'infrastructure aux fins de son optimisation (Murray, 2018, p. 201). Rappelons-nous aussi que la « morale artificielle » ne peut pas exister aujourd'hui. Il s'agit du programme qui agit à travers une interface. De fait, l'intelligence artificielle est susceptible d'être performante pour interpréter, prédire et manipuler le comportement humain au moyen de l'empathie. Cet accès sans

intermédiaire est susceptible de conférer à l'IA un avantage décisif dans de nombreux domaines, ainsi que le croit Nick Bostrom :

Je pense que le but ultime doit être la disparition totale du travail. Ce que nous voulons faire, c'est remplacer le besoin pour les humains de travailler, pas seulement grâce à l'intelligence artificielle mais grâce à l'automatisation dans son ensemble – si nous avons inventé la charrue, c'est pour arrêter de nous briser le dos. [...] Je ne suis pas convaincu qu'actuellement ces technologies d'intelligence artificielle aient un grand impact sur le marché du travail, je pense que tout cela est encore une possibilité dans le futur. Il faut être attentif à la différence entre la capacité des machines à effectuer certaines tâches et leur capacité à occuper certains emplois. [...] [D]ans le futur, la technologie ne servira plus seulement à transformer le monde extérieur, mais [...] elle offrira des possibilités de transformer la nature humaine en étendant nos capacités humaines (Bostrom, 2017).

Nous avons eu recours à la notion de singularité pour la comparer à son acception en informatique. La singularité, dans ce domaine, correspond à une convergence des nanotechnologies associées à la robotique, pensées par l'informatique et où il devient possible de tout construire à partir de l'atome (Kurzweil, 2019). En conséquence, et si la superintelligence des appareils est possible, comme l'écrit M. Shanahan : « Une responsabilité énorme nous incombe. Nous devons décider ce qu'il faut faire avec la technologie, non seulement pour le bien de l'humanité, mais aussi pour préparer un futur pour la conscience elle-même » (Murray, 2018).

Nous avons pu nous rendre compte que la visite guidée reste un outil solide pour encadrer par avance les usages d'un objet numérique. Cela peut être expliqué par la définition de l'activité de guide situationnelle, recensée *in situ*, qui évoque un espace de rencontre, du lien social, de la situation rhétorique, la dynamique du groupe et l'appropriation du patrimoine individuel et collectif. Tous ces éléments représentent la motivation des visiteurs qui ont recours aux guides humains pour se déplacer sur les territoires visités. Cette activité pourrait ainsi servir de limite à la navigation libre, affranchie de la présence de l'autre et dématérialisée. Ce cadre met en avant l'œuvre de l'esprit humain qui réunit deux conditions : revêtir une forme perceptible par les sens, être originale et tracer ainsi une ligne de partage entre créativité humaine et automatisation technique. Dans cette perspective, le travail de guide s'inscrit dans l'intérêt général. Située ainsi, son action prend place au carrefour entre secteur public et secteur privé. La réponse aux tensions issues d'une telle situation pourrait être la reconversion de la politique patrimoniale en politique publique. La politique

patrimoniale devenue publique ne viserait plus à réguler le guide selon les lois, mais à établir les normes que le guide serait amené à adopter de son propre gré. Le comportement imposé par des normes deviendrait comme inné. Il en découlerait une nécessité de transparence dans la transposition des normes. Mais de quel guide pourrait-on encore parler ?

RÉFÉRENCES

Instruments juridiques

Conseil de l'Europe

Convention de Faro sur la valeur du patrimoine culturel pour la société, Conseil de l'Europe, STCE n° 199, adoptée le 27 octobre 2005 (entrée en vigueur le 1^{er} juin 2011).

Union européenne

CE, *Directive 2013/55/UE du Parlement européen et du Conseil du 20 novembre 2013 modifiant la directive 2005/36/CE relative à la reconnaissance des qualifications professionnelles et le règlement (UE) n° 1024/2012 concernant la coopération administrative par l'intermédiaire du système d'information du marché intérieur (« règlement IMI »)*, [2013] JO, L 354/132.

CE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union projet de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle, dit « Artificial Intelligence Act »*, COM (2021) 206 final, 21 avril 2021.

Doctrine

Allard, Michel et Bernard Lefebvre (dir.), *La formation en muséologie et en éducation muséale à travers le monde*, Montréal, MultiMondes, 2001.

Appiotti, Sébastien, Marie Ballarini et Marie-Laure Bernon, « Les expositions en ligne », *Rencontres musée Île-de-France*, Paris, Association Mêtis, janvier-mars 2022, <<https://metis-lab.com/2021/11/17/rencontres-museo-idf-les-expositions-en-ligne/>>.

Bach, Jean-François, Olivier Houdé, Pierre Léna et Serge Tisseron, « L'Enfant et les écrans / Un avis de l'Académie des sciences », dans Stéphane Vial (dir.), *L'être et l'écran : comment le numérique change la perception*, Paris, Presses universitaires de France, 2013.

Balibar, Étienne, Séminaire « Le droit d'avoir les droits et la citoyenneté cosmopolite », Paris, Collège de France, 2013.

Barnet, Jamie, *L'homme ou la machine ? Comment Internet tue la démocratie (et comment la sauver)*, Paris, De Boeck Supérieur, 2019.

Bergeot, Patrick, Retour sur l'expérimentation « Livre blanc / Visite guidée à distance », Centre des monuments nationaux (CMN), *Rencontres musée Île-de-France*, Paris, Association Mêtis, janvier-mars 2022, <<https://metis-lab.com/2021/11/17/rencontres-museo-idf-les-expositions-en-ligne/>>.

Berns, Thomas, « Sortir de la répétition de la gouvernementalité algorithmique », dans Bernard Stiegler (dir.) *et al.*, *La toile que nous voulons*, Paris, FYP Éditions, 2017, p. 77-88.

- Bitzer, Lloyd F., « The Rhetorical Situation », *Philosophy & Rhetoric*, vol. 1, n° 1, January 1968, p. 1-14.
- Bonnaud-Le Roux, Samuel, « Les enjeux juridiques liés à la numérisation tridimensionnelle du patrimoine », *In Situ*, n° 42, 2020.
- Boucher, David, Diane Duquet et Nicole Beaudry, Commission de l'éthique de la science et de la technologie, Avis remis au ministre du Développement économique, de l'Innovation et de l'Exportation intitulé « Viser un juste équilibre : un regard éthique sur les nouvelles technologies de surveillance et de contrôle à des fins de sécurité », Québec, 2008.
- Boudon-Machuel, Marion et Barthélemy Serres, « SCULPTURE3D, du projet de laboratoire à l'exposition numérique », *In Situ*, n° 42, 2020.
- Bostrom Nick, « Entretien conduit par Georges Benoît », *Les Échos*, Paris, 26 juin 2017.
- Cardon Dominique, « Le Web que nous voulons en 8 proportions », dans Bernard Stiegler *et al.* (dir.), *La toile que nous voulons*, Paris, FYP Éditions, 2017, p. 23-44.
- Charland, Maurice, « Le langage politique » dans Anne-Marie Gingra (dir.), *La communication politique : état des savoirs, enjeux et perspectives*, Québec, Presses de l'Université du Québec, 2003, p. 67-91.
- Chaumier, Serge et François Mairesse, *La médiation culturelle*, Paris, Armand Colin, 2013.
- Cohen-Azria, Cora et Dominique Lahanier-Reuter (dir.), *Conscience disciplinaire : les représentations des disciplines à la fin de l'école primaire*, Rennes, Presses universitaires de Rennes, 2013.
- Corral-Regourd, Martine et Christophe Alcantara, « Ecosystème muséal : une notion encore vacillante », dans Serge Agostinelli et Nicole Koulayan (dir.), *Les écosystèmes numériques : intelligence collective, développement durable, interculturalité, transfert de connaissance*, Paris, Presses des Mines. Série « Design numérique », 2016, p. 59-70.
- Courrèges d'Agnos, Constance, Jean-François Dubos et Sylvie Legrosse, « Le numérique au service de l'exceptionnel à la bibliothèque du Service historique de la Défense : le rouleau de l'ordre de bataille de Vélez-Málaga », *In Situ*, n° 42, 2020.
- De Ganay, Claude et Dominique Gillot, sénatrice, au nom de l'Office parlementaire d'évaluation des choix scientifiques et technologiques n° 464, Tome I (2016-2017), 15 mars 2017, partie « C. La prise en compte grandissante des enjeux éthiques » <<https://www.senat.fr/rap/r16-464-1/r16-464-113.html>> (consulté en février 2022).
- Degrigny, Christian, Jean-Philippe Farrugia, Frédéric Mérienne et Matthieu Pinette, « Optimisation virtuelle d'un décor princier de la fin du xive siècle au château de Germolles (Saône-et-Loire) », *In Situ*, n° 42, 2020.
- Dewey, John, *Experience and Nature*, New York, Dover Publication, 1958.
- Eidelman, Jacqueline *et al.* (dir.), *Inventer des musées pour demain. Rapport de la mission musées XXI^e siècle*, Paris, La Documentation française, 2017.
- Falk, John et Lynn Derking, *The Museum Experience*, Washington (DC), Whalesback Books, 1994.
- Falk, John et Lynn Derking, *Learning from Museums. Visitor Experience and the Making of Meaning*, Plymouth, Altamira Press, 2000.
- Foucault, Michel « De gouvernement de soi et des autres », Paris, « Cours de Collège de France », 1^{er}-5 janvier 1983.

- Gellereau, Michèle, « Mémoire du travail, mémoire des conflits. Comment les témoignages se mettent en scène dans les visites patrimoniales », dans *Communication & Langages*, n° 149, 2006.
- Gellereau, Michèle, *Les mises en scène de la visite guidée*, Paris, L'Harmattan, 2004.
- Grenard, Agnes, « Normalisation, certification : quelques éléments de définition », dans *Revue d'économie industrielle*, vol. 75, 1^{er} trimestre 1996, p. 45-60.
- Guidi, Diletta, Magali Jenny et Alix Hagen, « Enquête sur les musées suisses à l'ère du digital », Fribourg, Groupe de recherche Musée 2.0, 2021.
- Holbrook, Morris et Elizabet Hirschman, « The experience aspects of consumption : Consumer fantasies, feelings, and fun », dans *Journal of Consumer Research*, vol. 9, n° 2, 1982, p. 132-140.
- Hui, Yuk, « Le concept de groupe dans les réseaux sociaux. Éléments pour une mécanologie de la participation », dans Bernard Stiegler *et al.* (dir.), *La toile que nous voulons*, Paris, FYP Éditions, 2017, p. 167-187.
- Jorion, Paul, « La mise à l'écart de l'homme par la machine est-elle irréversible ? », dans Bernard Stiegler *et al.* (dir.), *La toile que nous voulons*, Paris, FYP Éditions, 2017, p. 63-76.
- Kelsen, Hans, *Théorie générale des normes* (trad. Olivier Beaud et Fabrice Malkani), Paris, Presses universitaires de France, coll. Léviathan, 1996.
- Kurzweil, Ray, *Singularity is Near*, Londres, Penguin Books, 2019.
- Lebel, Estelle, « L'image politique », dans Anne-Marie Gingras (dir.), *La communication politique : état des savoirs, enjeux et perspectives*, Québec, Presses de l'Université du Québec, 2003, p. 97-135.
- Le Cun, Yann, *Quand la machine apprend : la révolution des neurones artificiels et de l'apprentissage profond*, Paris, Éditions Odile Jacob, coll. L'homme qui révolutionne l'intelligence artificielle, 2019.
- Liévaux, Pascal et Livio De Luca, « Imagerie numérique et patrimoine culturel : représentation et transmission des connaissances », *In Situ*, n° 42, 2020.
- Lingle Pond, Kathleen, *The Professional Guide : Dynamics of Tour Guiding*, New York/Chichester/Brisbane/Toronto/Singapore, John Wiley & Sons Inc., 1993.
- Mairesse, François et Fabrice Rochelandet, *Économie des arts et de la culture*, Paris, Armand Colin, 2015.
- Montpetit, Raymond, « Expositions, parcs, sites : des lieux d'expérience patrimoniales », *Culture et musées*, Actes Sud, n° 5, 2005, p. 111-132.
- Murray, Shanahan, *La singularité technologique. Intelligence artificielle, superintelligence et futur de l'humanité*, Paris, FYP Éditions, 2018.
- Renaud, Lise, Université d'Avignon, Centre Norbert Elias, « Exploration des formats médiatiques », *Rencontres muséo Île-de-France*, Paris, Association Métis, janvier-mars 2022, <<https://metis-lab.com/2021/11/17/rencontres-museo-idf-les-expositions-en-ligne/>>.
- Rouvroy, Antoinette, et Thomas Berns, « Gouvernamentalité algorithmique et perspectives d'émancipation / Le disparate comme condition d'individuation par la relation ? », *Réseaux*, vol. 2013/1, n° 177, p. 163-196.
- Sadin, Éric, *La vie algorithmique : critique de la raison numérique*, Paris, Édition L'échappée, 2005.
- Schmitt, Daniel et Muriel Meyer-Chemenska, « 20 ans de numérique dans les musées : entre monstration et effacement », *La Lettre de l'OCIM*, n° 162, 2015.

- Saemmer, Alexandra (Cemti, Université Paris 8) et Nolwenn Tréhondart (CREM, Université de Lorraine), « Livres d'art numériques, de la conception à la réception. Retour d'expérience sur une recherche collaborative », Intervention dans la journée d'étude *Les substituts numériques des expositions*, Paris, Maison de la recherche de la Sorbonne Nouvelle Paris 3, 4 février 2022.
- Sorin, Calude Cristian et Giuseppe Longo, « Le déluge des corrélations fallacieuses dans le big data », dans Bernard Stiegler *et al.* (dir.), *La toile que nous voulons*, Paris, FYP Éditions, 2017, p. 123-160.
- Stiegler, Bernard, Conférence « The flood of rights » / « La crue des droits », Arles, Luma Foundation, 17 septembre 2021, <<https://www.luma.org/fr/live/watch/La-crue-des-droits-Bernard-Stiegler-Eric-Kluitenberg-607ebf86-a794-4874-b284-3847440ffb81.html>> (consulté en février 2022).
- Stiegler, Bernard, « Ce que nous entendons par Web néguentropique », dans Bernard Stiegler (dir.) *et al.*, *La toile que nous voulons*, Paris, FYP Éditions, 2017, p. 10-22.
- Turinetti, Alice, *La normalisation : étude en droit économique*, Paris, Connaissances & savoirs, 2018.
- Uzlyte, Lina, *Évolution de la compétence de guide conférencier en Europe*, Thèse de doctorat, Université de la Sorbonne Nouvelle Paris 3, 2021.
- Vial, Stéphane, *L'être et l'écran : comment le numérique change la perception*, Paris, Presses universitaires de France, Paris, 2013.
- Volle, Michel, « Comprendre la croissance à l'heure de l'information de la société, dans Philippe Lemoine (dir.), *Une croissance intelligente*, Paris, Paris Descartes & Cie, 2012.

PARTIE IV

LES RÉPERCUSSIONS DE L'INTELLIGENCE ARTIFICIELLE SUR LA PROPRIÉTÉ INTELLECTUELLE, LES DROITS HUMAINS ET LA DIVERSITÉ CULTURELLE

IA ET CULTURE

DE LA CRÉATION AUX DONNÉES

Alexandra Bensamoun¹

Intelligence artificielle et culture ? Illustrations. Quoi de commun entre l'intelligence artificielle (IA) et la culture ? D'un côté, une des technologies les plus innovantes, promesse de progrès, mais aussi siège de risques nouveaux et de peurs ; de l'autre, l'expression artistique, synonyme de plaisir, de liberté, de diversité... L'IA a investi notre quotidien. Le secteur culturel n'y a pas échappé et cette intrusion revêt différents aspects, tant artistiques qu'économiques.

Les systèmes d'IA² s'enrichissent de plusieurs techniques. Si la médiatisation se concentre surtout sur les IA génératives ou systèmes auto-apprenants (apprentissage machine ou *machine learning*), ceux-ci coexistent avec les approches logiques (ou déterministes) et statistiques. Pour autant, la création culturelle est souvent associée à l'IA générative : le système fonctionne grâce à l'ingestion d'œuvres, décomposées et analysées, afin d'identifier des traits caractéristiques communs, des valeurs, des

1. Professeure de droit à l'Université Paris-Saclay.

2. L'expression est issue de : CE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union projet de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle*, dit « Artificial Intelligence Act », COM (2021) 206 final, 21 avril 2021, art. 3 (1) : « un logiciel qui est développé au moyen d'une ou plusieurs des techniques et approches énumérées à l'annexe I et qui peut, pour un ensemble donné d'objectifs définis par l'homme, générer des résultats tels que des contenus, des prédictions, des recommandations ou des décisions influençant les environnements avec lesquels il interagit ». La référence à l'annexe I, qui nuit à la neutralité technologique de la norme, devrait disparaître.

tendances. Ainsi, la réalisation du portrait *Edmond de Bellamy*, tableau vendu aux enchères chez Christie's plus de 400 000 dollars américains en 2018 et représentant un personnage imaginaire du XIX^e siècle, a été permise grâce à la constitution d'une base d'entraînement de près de 15 000 portraits classiques du XIV^e au XX^e siècle ; de même, *Daddy's Car*, musique composée à la manière des Beatles par Flow Machines de Sony, s'est inspiré d'un corpus de 44 créations du quatuor anglais.

Enjeux. À envisager le rapprochement de l'IA et de la culture, ces illustrations sont sans doute les plus emblématiques, par leur médiatisation. Le rattachement exclusif à l'IA est pourtant très exagéré, car dans les deux hypothèses, la personne humaine est en réalité essentielle dans le processus créatif. Surtout, ces réalisations artistiques ne sont que le sommet émergé de l'iceberg. En effet, l'IA ne s'est pas immiscée que dans les aspects créatifs. Le champ est étroitement corrélé à celui des données, permettant des applications concrètes qui occupent toute la chaîne de valeur, du stade de la création à celui de la consommation, en passant par la production. Les données nourrissent l'IA, l'informent, l'orientent, dans un lien inextricable : elles sont nécessaires à son fonctionnement et sont, de ce point de vue, devenues le véritable enjeu de l'émergence de ces technologies (Bruguière, 2020).

Les thématiques sont ainsi liées et l'Europe l'a bien compris. De la société de l'information au marché unique numérique, l'Union européenne a entrepris d'appliquer à la propriété intellectuelle sa nouvelle stratégie, notamment fondée sur la donnée (Communications de la Commission européenne, 2020a ; 2020b) et sur les potentialités associées en matière d'IA³. La France contribue à la réflexion, notamment grâce aux rapports élaborés dans le cadre du Conseil supérieur de la propriété littéraire et artistique (CSPLA), organe de conseil du ministre de la Culture (Bensamoun, Farchy et Schira, 2020 ; Bensamoun et Bouquerel, 2020⁴). La première ministre a également missionné en septembre 2023 des experts

3. CE, *Résolution du Parlement européen du 20 octobre 2020 sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle*, 2020/2015(INI), [2020] JO, C 404/107.

4. Certaines des réflexions qui suivent sont issues de rapports au CSPLA, notamment sur l'IA et la culture (A. Bensamoun, J. Farchy et P.-F. Schira, « Les enjeux juridiques et économiques de l'intelligence artificielle dans les secteurs de la création culturelle », janv. 2020) et sur la fouille de textes et de données (A. Bensamoun et Y. Bouquerel, « Transposition des exceptions de fouille de textes et de données », déc. 2020). Les rapports remis par les personnalités qualifiées sont disponibles sur le site du ministère de la Culture. Ministère de la Culture, « Missions du CSPLA », <<https://www.culture.gouv.fr/fr/Nous-connaître/Organisation-du-ministère/>>

internationaux, au sein du Comité de l'intelligence artificielle, avec l'espoir d'éclairer les politiques sur la voie à emprunter⁵. Mais le sujet est loin d'être épuisé. Et la réflexion est d'ailleurs toujours pendante au niveau européen (European Commission, 2022a ; 2022b).

Annnonce. Phénomène conquérant, l'IA s'immisce ainsi dans la culture au point d'imposer des réflexions nouvelles. En ce sens, elle implique de s'interroger sur la création culturelle, précisément sur la place des productions algorithmiques, générées par une IA (1). En outre, assise sur les données, elle s'invite dans toute la chaîne de valeurs et redessine les équilibres au sein des industries culturelles (2).

1. IA ET CRÉATION CULTURELLE

IA et création : annonce. Les productions culturelles de l'IA ressemblent à s'y méprendre aux créations humaines. Dès lors, le droit d'auteur semble en être le réceptacle naturel (Vivant, 2018 ; Bensamoun, 2022, p. 285 ; Binctin, 2020, p. 41⁶). Encore faut-il que ces productions répondent aux conditions de qualification de l'œuvre de l'esprit, objet du droit d'auteur (1.1). Et même en concédant ce rapprochement, le régime existant nécessitera peut-être des aménagements (1.2).

1.1 La qualification des productions de l'IA

Les hypothèses exclues. À ce stade, il convient de souligner que le raisonnement se limitera aux véritables productions algorithmiques, c'est-à-dire *générées* par une IA, délaissant en conséquence les créations *assistées* par une IA. En effet, pour ces dernières, l'IA reste un outil – certes très perfectionné – au service de la création humaine. Dans ce cadre, le droit d'auteur a classiquement vocation à s'appliquer sans obstacle. C'est le cas notamment pour les exemples cités en introduction. Dans ces hypothèses, on peut identifier de véritables auteurs : en dépit des annonces médiatiques ambiguës (la toile intitulée *Edmond de Bellamy* est même signée d'une fonction mathématique, par référence à l'algorithme utilisé), les créations ont

Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA>.

5. Gouvernement de la République française, « Comité de l'intelligence artificielle générative », <<https://www.gouvernement.fr/communique/comite-de-lintelligence-artificielle>>.
 6. Voir aussi CE, *Résolution du Parlement européen du 20 octobre 2020 sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle*, 2020/2015(INI), [2020] JO, C 404/107.

bien été réalisées par des personnes physiques qui ont marqué l’œuvre de leur personnalité en intervenant de manière originale dans l’univers des formes – c’est là la condition même de la qualification d’œuvre. Même technologique, l’IA demeure alors un instrument de la création, utilisé par l’auteur-créateur. Au rebours, les productions générées par une IA sont censées se passer de créateur. La personne physique se métamorphose alors en simple technicien ou exécutant, sans possibilité d’exprimer son for intérieur dans la création.

Une autre hypothèse doit également être écartée : celle d’une reconnaissance de l’atorat de l’IA sur les créations algorithmiques (Yanisky-Ravid et Velez-Hernandez, 2018 ; Pearlman, 2018). L’exclusion s’appuie sur deux arguments. Le premier, relatif à la technique technicienne, prend acte de l’état de la technique : l’autonomie de l’IA créative est en réalité très largement orientée par des décisions humaines concernant le choix de la base d’apprentissage, du protocole d’apprentissage, de l’architecture interne du modèle, etc. De manière générale, les capacités de l’IA sont souvent fantasmées : l’IA forte, consciente, n’existe pas. Le second argument repose sur la technique juridique : l’IA ne saurait se voir reconnaître une personnalité juridique, au risque de créer une véritable chimère, mi-objet, mi-sujet de droit (Loiseau, 2018), et conduisant *in fine* à déresponsabiliser les acteurs concernés (Bensamoun et Loiseau, 2017, p. 2063). Si pareille solution a un temps pu être évoquée par le Parlement européen⁷, elle a ensuite été largement rejetée par toutes les autorités intergouvernementales.

Les conditions de la protection. Le rattachement des créations algorithmiques à la catégorie des œuvres de l’esprit implique de les passer au crible de la qualification (Bensamoun, Farchy et Schira, 2020 ; Larrieu, 2014, p. 465 ; Michaux, 2018, p. 40 ; Bensamoun, 2020a, p. 75). L’œuvre de l’esprit est généralement définie comme une création de forme originale. Au sens du droit européen – et aux termes d’une jurisprudence nourrie (Benabou, 2014, p. 17⁸) –, la forme implique que l’objet soit identifiable avec

-
7. CE, *Résolution du Parlement européen du 16 févr. 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique, 2015/2013(INL)*, [2017] JO, C 252/239, au para 59 f), qui proposait de réfléchir à « [...] la création, à terme, d’une personnalité juridique spécifique aux robots, pour qu’au moins les robots autonomes les plus sophistiqués puissent être considérés comme des personnes électroniques responsables, tenues de réparer tout dommage causé à un tiers », ajoutant qu’« il serait envisageable de conférer la personnalité électronique à tout robot qui prend des décisions autonomes ou qui interagit de manière indépendante avec des tiers ».
 8. Et voir plus récemment, CJUE, grde ch., *Levola Hengelo*, aff. C-310/17, 13 nov. 2018 ; CJUE, *Cofemel*, aff. C-683/17, 12 sept. 2019 ; CJUE, *Brompton*, aff. C-833/18, 11 juin 2020.

suffisamment de précision et d'objectivité. En outre, la reconnaissance d'une originalité suppose l'expression par l'auteur de choix libres et créatifs.

La première condition ne semble pas poser de difficultés dans la mesure où les créations algorithmiques s'expriment bien par une forme clairement extériorisée. La seconde, relative à l'originalité, suscite davantage de doutes. D'abord, elle exige une certaine créativité, une liberté créative, qui pourrait faire défaut en raison du caractère contraint de la création algorithmique. Cependant, en réalité, le même modèle algorithmique peut donner naissance à de multiples créations ; précisément, l'apprentissage peut apparaître comme une technique d'émancipation qui permet de reconnaître cette liberté créative. Ensuite, la condition d'originalité implique la présence d'un auteur personne physique, celui-là même qui exprime sa personnalité dans l'œuvre. Or, son identification peut s'avérer compliquée en matière de production algorithmique. La démarche serait néanmoins possible si l'on accepte, d'une part, de reconnaître l'originalité *dans l'œuvre* et non plus dans son auteur et, d'autre part, d'identifier un auteur plus indirect, c'est-à-dire plus éloigné de la création (Bensamoun, 2020b, p. 13). Certes, l'utilisateur, qui n'opère aucun choix créatif et se contente de « commander » la création, doit être exclu de la qualité d'auteur (Ginsburg, 2018, p. 131) : il ne bénéficie que de l'aléa, du hasard, non de l'arbitraire nécessaire à l'expression du libre arbitre (Binctin, 2020, p. 57 et s.). Mais il ne saurait en être de même dans l'hypothèse où il agit sur la création en y imprimant ses propres choix créatifs, notamment en amont, dans la sélection de la base d'entraînement, et en aval, au stade de la curation (Deltorn, 2016, p. 4). C'est ici aussi une question de frontière et surtout de qualification. À quel moment la personne physique qui intervient marque la réalisation de sa personnalité ? Quel degré d'intervention est-il requis ? En quels endroits et de quelle manière ? En outre, le concepteur de l'IA pourrait, selon certains, être qualifié d'auteur ; il enserre en effet la production algorithmique dans un champ des possibles qui le met en lien avec la création. À ce titre, il imprime ses choix dans l'outil, donc indirectement dans la création algorithmique. Cependant, les usages actuels de l'IA générative, qui s'accommode désormais de techniques d'optimisation (comme le *fine-tuning*), rend le raisonnement difficile à maintenir.

La nécessité d'un auteur. La production générée par une IA semble donc échouer au stade de la qualification du fait de l'absence d'auteur personne physique qui marquerait l'œuvre de sa personnalité par ses choix libres et

créatifs. C'est d'ailleurs la position aujourd'hui soutenue aux États-Unis par l'US Copyright Office⁹ et la jurisprudence¹⁰.

1.2 Le régime des productions de l'IA

Un droit d'auteur adapté ? Au-delà de la parenté matérielle entre la production algorithmique et la création humaine, et à considérer que la qualification est acquise (ce qui n'est pas simple), le droit d'auteur pourrait recevoir les productions de l'IA. Le régime du droit d'auteur aurait alors pleinement vocation à s'appliquer. Cependant, si, à l'avenir, une insuffisance du droit positif était rapportée, une intervention législative pourrait être nécessaire.

Reste à savoir si une adaptation des règles de droit commun du droit d'auteur serait pertinente. L'ampleur des prérogatives reconnues en droit d'auteur pourrait sembler inadaptée aux créations algorithmiques. Aussi, comme en matière de logiciel par exemple, il pourrait être proposé d'ajuster la portée des prérogatives, les règles de titularité ou encore les exceptions. Notamment, dans la mesure où le lien à l'auteur paraît plus distendu, le droit moral pourrait être amoindri.

Les régimes existants. Aucun texte européen ou international n'envisage pour l'heure la protection des productions algorithmiques. La réflexion est en cours au niveau européen (Communication de la Commission européenne, 2020a). Cependant, certains États les ont déjà incorporées dans leur droit d'auteur interne pour leur offrir un régime *ad hoc*. Notamment, le *Copyright, Designs and Patents Act* anglais isole les « œuvres générées par ordinateur », définies comme celles où il n'y a pas d'auteur humain¹¹. Dans cette hypothèse précise, le texte pose une fiction : l'auteur n'est pas le créateur, mais la personne qui a pris les « dispositions nécessaires » pour la création¹², ce qui place cette désignation sous la dépendance de l'interprétation judiciaire, l'expression ne permettant pas d'identifier clairement le bénéficiaire. En outre, la durée de la protection patrimoniale est réduite à cinquante ans, au lieu des soixante-dix ans de principe, et le point de

9. Voir notamment *Le Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*, 16190 Federal Register, vol. 88, no. 51, 16 mars 2023.

10. US District Court for the District of Columbia, Case n° 1:22-cv-01564, *Thaler v. Perlmutter*, 18 août 2023.

11. *Copyright, Designs and Patents Act* (CDPA, R.-U.), 1988, art. 178 (b).

12. *Ibid.*, ch. 48, art. 9 (3).

départ est la réalisation de l'œuvre et non la mort de l'auteur¹³. S'agissant du droit moral, il est fait exception au droit à la paternité¹⁴ et au droit au respect de l'intégrité de l'œuvre¹⁵. La Nouvelle-Zélande¹⁶ et l'Irlande¹⁷ connaissent des systèmes similaires.

Un nouveau droit ? Il est aussi envisageable de protéger les productions de l'IA par un droit spécial, *sui generis*, peut-être fondé sur l'investissement, comme l'est le droit des producteurs de bases de données en Europe. L'air du temps n'est cependant pas à la création de nouveaux droits spéciaux objets de réservation.

L'absence de réservation. Au contraire, certains membres de la doctrine ont pu proposer une application négative du droit d'auteur, considérant que ces œuvres sans auteur devraient être qualifiées de « communs par nature » – *commons by design* – et intégrer, à ce titre, un domaine public par défaut (Azzaria, 2018, p. 925 ; Cabay, 2019, p. 307). Cette qualification extensive du domaine public reposerait sur deux conditions : d'une part, que l'IA joue un rôle prépondérant dans l'expression de l'œuvre et, d'autre part, que l'intervention humaine soit en conséquence résiduelle, ne permettant pas d'imprimer des choix créatifs. L'absence de protection a pu également être soutenue pour des raisons écologiques (coût écologique des productions numériques de masse) et en considération du risque d'appauvrissement culturel.

Cependant, l'exclusion de tout droit privatif présente également des désavantages. D'abord, elle laisse sans protection les candidats à la titularité de droits voisins, précisément les artistes-interprètes de productions algorithmiques. En effet, l'attribution de droits est soumise à la qualification en amont d'œuvre de l'esprit. Ensuite, cette voie favoriserait les comportements de contournement dans la mesure où le « producteur » de la réalisation algorithmique n'aurait aucune incitation à

13. *Ibid.*, art. 12 (7).

14. *Ibid.*, art. 79 (2) (a).

15. *Ibid.*, art. 81 (2).

16. Art. 5 (2) (a) du Copyright Act, 1994 : « For the purposes of subsection (1), the person who creates a work shall be taken to be, – (a) in the case of a literary, dramatic, musical, or artistic work that is computer-generated, the person by whom the arrangements necessary for the creation of the work are undertaken ».

17. Art. 21 (f) du Copyright and Related Rights Act, 2000 (n° 28/2000) : « In this Act, "author" means the person who creates a work and includes : (f) in the case of a work which is computer-generated, the person by whom the arrangements necessary for the creation of the work are undertaken ».

révéler que celle-ci est générée par une IA. En taisant le processus créatif, il obtiendrait le bénéfice du droit d’auteur classique.

Un choix éthique et politique. En définitive, le choix d’un statut des productions culturelles de l’IA apparaît comme une question éthique, s’intégrant plus largement à la réflexion sur la place de l’IA dans notre société. Il requiert également sans doute un acte politique. Il appartiendra ainsi, en dernier état, au législateur de trancher pour dégager la voie privilégiée.

2. IA ET DONNÉES DE LA CULTURE

IA et données : annonce. L’IA a fait émerger un nouveau problème transversal : celui des données. Dans le domaine de la culture, la question est double, soit que l’œuvre puisse elle-même être qualifiée de donnée (2.1), soit que les données portent sur l’œuvre (2.2).

2.1 L’œuvre comme donnée

La fouille de textes et de données. Si la réflexion concernant le statut des productions générées par une IA est pendante, celle sur l’amont de la création est plus aboutie. Utilisant les techniques d’apprentissage-machine, les outils d’IA générative sont nourris par des masses de données. Par ailleurs, au-delà de ces techniques particulières, l’opération même d’extraction de données est déjà, pour certains, une technique d’IA.

Or, parmi les données analysées, se trouvent des œuvres protégées par le droit d’auteur. Se pose alors la question de savoir si la fouille de données-œuvres, opérée en vue de réaliser une création algorithmique, met en œuvre le monopole ou non. Considérant que le droit de lire implique le droit de fouiller (*the right to read is the right to mine*), certains acteurs avancent que la fouille de textes et de données (appelée en anglais *text and data mining* ou TDM) était, par construction, licite dès lors que l’accès à l’œuvre est lui-même licite. Ils soutiennent que le droit d’auteur ne protège que l’originalité de l’œuvre initiale, laquelle se trouve totalement déconstruite dans un projet de fouille. Pour d’autres, les contenus protégés n’étant pas utilisés pour eux-mêmes, mais pour leur contenu informationnel, aucun acte d’exploitation ne peut être caractérisé. En outre, ces copies ne génèrent pas d’acte d’exploitation distinct du processus technique (Dussolier, 2014, p. 267). Pour autant, en pratique, les fouilles impliquent des reproductions d’œuvres protégées, lesquelles sont, sauf exception, soumises au droit d’auteur, peu important l’usage fait de ces reproductions.

Les exceptions européennes. L'incertitude pesant sur cette situation et la volonté européenne de permettre le développement de ces technologies ont poussé le législateur européen à réglementer la situation dans le sens d'une libéralisation des usages. Ainsi, la directive 2019/790 du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique¹⁸ a prévu deux dispositifs en faveur de la fouille de textes et de données.

Le texte européen définit d'abord la fouille de textes et de données comme « toute technique d'analyse automatisée visant à analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations¹⁹ ». Il prévoit ensuite deux exceptions ou limitations obligatoires au droit de reproduction, que les États membres ont donc l'obligation de transposer. Le bénéfice de ces dispositions dérogatoires est conditionné, dans les deux cas, à un accès licite aux contenus protégés²⁰, incluant la mise à disposition libre en ligne.

La première exception²¹, à finalité de recherche scientifique, permet la fouille de tout type de contenus protégés, par les organismes de recherche et les institutions du patrimoine culturel. Elle ne saurait être remise en cause ni par la voie contractuelle, ni par l'exercice du droit de retrait permis dans la seconde hypothèse dérogatoire.

La seconde exception²² diffère par sa très large portée, puisqu'elle bénéficie à tous les acteurs, quelle que soit la finalité poursuivie. Cependant, les titulaires de droits jouissent en contrepartie d'une option de retrait (*opt out*), en vertu de laquelle ils ont le droit, en utilisant des procédés lisibles par la machine²³, de s'opposer à la fouille, impliquant alors un retour à l'exclusivité. De nombreux titulaires européens (par exemple en France la SACEM) tentent d'exercer leur droit d'opposition, avec plus ou moins de succès. Quelles techniques seraient efficaces ? Comment assurer leur effectivité sans une transparence des acteurs sur les contenus fouillés ?

18. CE, *Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE*, [2019] JO, L 130/92.

19. *Ibid.*, art. 1^{er}.

20. Sur la notion d'accès licite, voir *ibid.*, le cons. 14.

21. *Ibid.*, art. 3.

22. *Ibid.*, art. 4.

23. Par ex. des métadonnées. Voir sur le sujet 2.2.

L'usage loyal, le *fair use* américain. Par comparaison, aux États-Unis, la fouille de textes de données semble principalement autorisée en vertu de la doctrine du *fair use*²⁴. Les tenants de cette analyse s'appuient sur l'arrêt *Google Books*, du 16 octobre 2015²⁵, dans lequel le juge a considéré que le fait de numériser plus de mille livres dans une bibliothèque sans l'autorisation des auteurs n'était pas contraire à l'usage loyal. En effet, la réalisation par Google de courts entrefilets contribue au caractère « hautement transformatif » de la fouille, permettant à l'utilisateur de déterminer si l'ouvrage en question correspond bien à ses recherches. Certains auteurs soutiennent cependant que la décision est très liée aux faits de l'espèce et qu'elle doit être considérée comme *sui generis* (Ginsburg, 2016). Au demeurant, l'hypothèse trouve en pratique une solution contractuelle dans la plupart des cas. De nombreux procès sont en cours aux États-Unis pour trancher cette question juridique de l'amont. Les titulaires de droits ont agi contre les producteurs d'IA générative pour s'opposer à l'utilisation de leurs contenus pour créer le modèle.

2.2 Les données sur l'œuvre

Diversité des données et des enjeux. Les données sur l'œuvre recouvrent diverses réalités. Elles sont aujourd'hui au cœur du fonctionnement des industries culturelles.

D'abord, certaines données correspondent à des pratiques professionnelles. Notamment, les données d'usage permettent de relier *un* utilisateur à *une* œuvre. Ces données relatives à l'utilisation des contenus culturels sont devenues un enjeu déterminant permettant notamment d'analyser les goûts des usagers, de définir leur profil et de leur proposer une offre personnalisée.

Ensuite, certaines données, qualifiées de métadonnées, fournissent des informations sur un contenu, un créateur ou toute autre caractéristique objective relative à l'œuvre. Elles peuvent être de nature juridique (identification des titulaires de droits), descriptive (titre, durée) ou encore enrichie (genre, analyse sémantique, rythme...).

Ces données permettent notamment, dans un univers d'abondance informationnelle, de recommander des contenus aux utilisateurs, par

24. Section 107 du *Copyright Act*, 17 USC § 107.

25. United States Court of Appeals, 2d Cir. 2015, *Authors Guild v. Google, Inc.*, n° 13-4829 (2d Cir. 2015).

rapprochement ou par comparaison. Les recommandations personnalisées sont le résultat de traitements automatisés algorithmiques, mais elles nécessitent souvent en amont, pour être pertinentes, des interventions humaines de qualification des contenus. Les données sur l'œuvre représentent dès lors un enjeu économique de taille pour les industries culturelles. Certes, on a pu dénoncer le risque d'enfermement du consommateur dans ses habitudes, dit « bulle de filtre ». En réalité, les algorithmes de recommandation peuvent également être utilisés pour promouvoir une diversité des contenus et leur « découvrabilité » (Mission franco-québécoise, 2020). Encore faut-il pour cela s'assurer de la transparence de ces démarches.

Les métadonnées (Azzi et El Hage, 2021) permettent également de circonscrire le champ des autorisations – donc de limiter les usages non autorisés –, de faciliter la conclusion de licences ou d'assurer une meilleure transparence de la rémunération des créateurs. En droit européen, elles bénéficient d'une protection juridique en tant que mesures techniques d'information²⁶, contre la suppression ou la modification lorsque l'information porte sur le régime des droits. Cette protection s'étend à la distribution, l'importation, la radiodiffusion, la communication au public ou la mise à disposition de contenus protégés dont les informations sur le régime des droits ont été supprimées ou modifiées sans autorisation.

Les enjeux économiques liés aux données de la culture, comme de tous les secteurs, sont de plus en plus importants. Se pose dès lors la question de l'accès et du partage des données²⁷.

Données et plateformes contributives. Les métadonnées sont aussi au cœur du dispositif mis en place par le fameux article 17 de la directive sur le droit d'auteur dans le marché unique numérique (Pollaud-Dulian, 2021, p. 77 ; Vivant, 2019 ; Sirinelli, 2019, p. 288 ; Alleaume, 2019, p. 530 ; Bensamoun, 2019, p. 89 ; Bensamoun, 2021, p. 399²⁸). Ce nouveau régime permet de rendre opposables le droit d'auteur et les droits voisins aux

26. CE, Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information, [2001] JO, L 167/10, art. 7.

27. Voir en ce sens les propositions du CE, « Data Act » : Proposition de Règlement du Parlement européen et du Conseil fixant des règles harmonisées pour l'équité de l'accès aux données et de l'utilisation des données (règlement sur les données), COM (2022) 68 final, 23 février 2022.

28. CE, Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, [2019] JO, L 130/92, art. 17.

plateformes communautaires. En effet, les fournisseurs de services de partage de contenus en ligne sont réputés réaliser des actes de communication au public lorsqu'ils permettent la diffusion de contenus culturels protégés téléversés par les internautes. En outre, ils ne bénéficient plus désormais du régime de responsabilité conditionnée (dit *safe harbor*), accordé aux hébergeurs en vertu de la directive « e-commerce²⁹ », dans la mesure où ils ont un rôle actif tenant à l'organisation et à la promotion des contenus. De ce fait, sauf à engager leur responsabilité, ils doivent obtenir une autorisation des titulaires de droits pour ces contenus, autorisation qui se traduira souvent par une licence. Cependant, ils peuvent s'exonérer de leur responsabilité s'ils apportent la preuve qu'ils ont fourni leurs meilleurs efforts pour obtenir une autorisation, pour garantir l'indisponibilité des contenus signalés par les titulaires de droits, c'est-à-dire pour lesquels ceux-ci leur ont fourni « les informations pertinentes et nécessaires », et pour empêcher la réapparition de ces contenus. Les informations visées, qui garantiront l'indisponibilité et empêcheront un nouveau téléversement du même contenu, peuvent se concrétiser dans des métadonnées, traitées par des outils de reconnaissance de contenus (Mochon *et al.*, 2020; Mochon *et al.*, 2021).

IA et culture : propos conclusifs. Le croisement entre l'IA et la culture est réel, contemporain, même si les réponses sont encore balbutiantes. Il appartiendra sans doute au pouvoir politique de trancher : faut-il protéger les productions algorithmiques ? Le cas échéant, par le droit d'auteur, un droit d'auteur spécial, ou par un droit *sui generis* ? Doit-on permettre une meilleure circulation des données en obligeant à leur partage ? Plus largement, l'art et la culture sont-ils solubles dans l'IA ? L'IA est-elle un danger pour la culture ? « Ceci n'est pas une... œuvre ». L'IA deviendra ce que l'être humain en fera. Loin des fantasmes, la technologie peut être un formidable levier pour la culture. Emparons-nous-en !

29. CE, Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« directive sur le commerce électronique »), [2000] JO, L L 178/1, art. 14.

RÉFÉRENCES

Instruments juridiques

Législation européenne

CE, *Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« directive sur le commerce électronique »)*, [2000] JO, L 178/1.

CE, *Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information*, [2001] JO, L 167/10.

CE, *Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique*, 2015/2013(INL), [2017] JO, C 252/239.

CE, *Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE*, [2019] JO, L 130/92.

CE, *Résolution du Parlement européen du 20 octobre 2020 sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle*, 2020/2015(INI), [2020] JO, C 404/107.

CE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union projet de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle, dit « Artificial Intelligence Act »*, COM (2021) 206 final, 21 avril 2021.

CE, *« Data Act » : Proposition de Règlement du Parlement européen et du Conseil fixant des règles harmonisées pour l'équité de l'accès aux données et de l'utilisation des données (règlement sur les données)*, COM (2022) 68 final, 23 février 2022.

Législation nationale

États-Unis

Copyright Act, 17 USC § 107.

Copyright Registration Guidance : Works Containing Material Generated by Artificial Intelligence, 16190 Federal Register, vol. 88, no. 51, 16 mars 2023.

Irlande

Copyright and Related Rights Act, 2000 (n° 28/2000).

Nouvelle-Zélande

Copyright Act, 1994.

Royaume-Uni

Copyright, Designs and Patents Act (CDPA, R.-U.), 1988.

Jurisprudence

Cour de justice de l'Union européenne (CJUE)

CJUE, grde ch., *Levola Hengelo*, aff. C-310/17, 13 nov. 2018.

CJUE, *Cofemel*, aff. C-683/17, 12 sept. 2019.

CJUE, *Brompton*, aff. C-833/18, 11 juin 2020.

Jurisprudence nationale

United States Court of Appeals, 2d Cir. 2015, *Authors Guild v. Google, Inc.*, n° 13-4829 (2d Cir. 2015).

US District Court for the District of Columbia, Case n° 1:22-cv-01564, *Thaler v. Perlmutter*, 18 août 2023.

Doctrine

Alleaume, Christophe, « L'article 17 de la directive [UE] 2019/790 : une [fragile] responsabilité des fournisseurs de service de partage en ligne de contenus protégés par un droit d'auteur ou un droit voisin », *Légipresse*, 2019.

Azzaria, George, « Intelligence artificielle et droit d'auteur : l'hypothèse d'un domaine public par défaut », *Les Cahiers de PI*, vol. 30, n° 3, 2018.

Azzi, Tristan et Yves El Hage, « Les métadonnées liées aux images fixes en ligne », *Rapport CSPLA*, juillet 2021.

Benabou, Valérie-Laure, « L'originalité, un Janus juridique. Regards sur la naissance d'une notion autonome de droit de l'Union », dans *Mélanges en l'honneur du professeur André Lucas*, Lexis Nexis, 2014.

Bensamoun, Alexandra et Grégoire Loiseau, « La gestion des risques de l'intelligence artificielle. De l'éthique à la responsabilité », *La Semaine juridique*, Édition générale, 2017, p. 2063-2072.

Bensamoun, Alexandra, « L'article 17 de la directive "Digital Single Market" ou comment la légitimité d'un droit se pare des atours de la valse... », *Légipresse*, vol. 2, n° 62, 2019.

Bensamoun, Alexandra, « Les créations générées par une intelligence artificielle ou la tentation du droit d'auteur », dans A. Bensamoun et F. Labarthe (dir.), *Culture et numérique : rencontre franco-québécoise*, Mare & Martin, 2020a.

Bensamoun, Alexandra, « Libres propos sur l'existence d'un droit de l'œuvre applicable aux créations issues de l'intelligence artificielle », dans *Penser le droit de la pensée : mélanges en l'honneur de Michel Vivant*, Paris, Dalloz, 2020b.

Bensamoun, Alexandra, Joëlle Farchy et Paul-François Schira, « Les enjeux juridiques et économiques de l'intelligence artificielle dans les secteurs de la création culturelle », janvier 2020, <<https://www.culture.gouv.fr/fr/Nous-connaître/Organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA>>.

Bensamoun, Alexandra et Yohann Bouquerel, « Transposition des exceptions de fouille de textes et de données », décembre 2020, <<https://www.culture.gouv.fr/fr/Nous-connaître/Organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA>>.

- Bensamoun, Alexandra, « Filtrage des contenus par les plateformes contributives et liberté d'expression », *Légipresse*, 2021.
- Bensamoun, Alexandra, « Propriété intellectuelle et intelligence artificielle », dans *Droit de l'intelligence artificielle*, A. Bensamoun et G. Loiseau (dir.), Lextenso, coll. Les Intégrales, 2^e éd., 2022.
- Binctin, Nicolas, « L'influence de l'intelligence artificielle sur les mécanismes de propriété intellectuelle », dans *Penser le droit de la pensée : mélanges en l'honneur de Michel Vivant*, Dalloz/LexisNexis, 2020.
- Bruguière, Jean-Michel, « Intelligence artificielle et droit d'auteur. Sortir de la science-fiction des "machines/auteurs", entrer dans la réalité du droit des données », *Comm. com. électr.*, Étude 11, juin 2020.
- Cabay, Julien, « Comparaison n'est pas raison », *Entertainment. Droit – Médias – Art – Culture*, 2019/6.
- Communication de la Commission européenne, « Exploiter au mieux le potentiel d'innovation de l'Union européenne. Un plan d'action en faveur de la propriété intellectuelle afin de soutenir la reprise et la résilience dans l'Union européenne », 25 nov. 2020, COM (2020) 760 final, 2020a.
- Communication de la Commission européenne, « Une stratégie européenne pour les données », COM (2020) 0066, 2020b.
- Communications de la Commission européenne, « Une nouvelle stratégie industrielle pour l'Europe », COM (2020) 0102, 2020c.
- Deltorn, Jean-Marc, « Droit d'auteur et créations des algorithmes d'apprentissage », *Propriétés intellectuelles*, n° 58, 2016.
- Dussolier, Séverine, « L'exploitation des œuvres : une notion centrale en droit d'auteur », dans *Mélanges en l'honneur du professeur André Lucas*, LexisNexis, 2014.
- European Commission, Directorate-General for Communications Networks, « Content and technology, study on copyright and new technologies : Copyright data management and artificial intelligence », 2022a, <<https://data.europa.eu/doi/10.2759/570559>>.
- European Commission, Directorate-General for Communications Networks, Content and Technology, K. Izsak, A. Terrier, S. Kreutzer *et al.*, « Opportunities and challenges of artificial intelligence technologies for the cultural and creative sectors », 2022b, <<https://data.europa.eu/doi/10.2759/144212>>.
- Ginsburg, Jane C., « People not machines : Authorship and what it means in the Berne Convention », *International Review of Intellectual Property and Competition Law*, vol. 49, n° 2, 2018.
- Ginsburg, Jane C., « Secure failure fair use analysis », *The Media Institute*, janvier 2016, <<https://www.mediainstitute.org/2016/01/25/security-failure-fair-use-analysis/>>.
- Gouvernement de la République française, « Comité de l'intelligence artificielle générative », 20 septembre 2023, <<https://www.gouvernement.fr/communiquer/comite-de-lintelligence-artificielle>>.
- Larriue, Jacques, « Le robot et le droit d'auteur », dans *Mélanges en l'honneur du professeur André Lucas*, New York, LexisNexis, 2014.
- Loiseau, Grégoire, « La personnalité juridique des robots : une monstruosité juridique », *La Semaine juridique*, n° 22, 2018, p. 1039-1042.

- Michaux, Benoît, « Singularité technologique, singularité humaine et droit d'auteur », dans *Droit, normes et libertés dans le cybermonde*, Mélanges Y. Poulet, Larcier, 2018.
- Mission franco-qubécoise sur la découvrabilité des contenus culturels francophones*, Rapport du ministère de la Culture et des Communications du Québec et ministère de la Culture de France, 2020.
- Mochon, Jean-Philippe et Alexis Goin, « Les outils de reconnaissance de contenus sur les plateformes de partage : propositions pour la mise en œuvre de l'article 17 de la directive sur le droit d'auteur dans le marché unique numérique », janvier 2021.
- Mochon, Jean-Philippe et Sylvain Humbert, « Vers une application effective du droit d'auteur sur les plateformes numériques de partage : état de l'art et propositions sur les outils de reconnaissance des contenus », janvier 2020.
- Pearlman, Russ, « Recognizing artificial intelligence (AI) as authors and inventors under U.S. intellectual property law », *Rich. J. L. & Tech.*, vol. 24, n° 2, 2018.
- Pollaud-Dulian, Frédéric, « Téléversement et responsabilité des prestataires de services de l'Internet : encore et toujours l'article 17 de la directive n° 2019/790 », *RTD com.*, 2021.
- Sirinelli, Pierre, « Le nouveau régime applicable aux fournisseurs de services de partage de contenus en ligne », *D. IP/IT*, mai 2019.
- Vivant, Michel, « Intelligence artificielle et propriété intellectuelle », *Comm. com. électr.*, Études 18, novembre 2018.
- Vivant, Michel, « Une responsabilité ad hoc pour les sites de partage (Commentaire de l'article 17 de la directive) », *Comm. com. électr.*, dossier 8, octobre 2019.
- Yanisky-Ravid, Shlomit et Luis Antonio Velez-Hernandez, « Copyrightability of artworks produced by creative robots and originality: A formality-objective model », *Minn. J.L. Sci. & Tech.*, vol. 19, n° 1, 2018.

BETWEEN DEVICE AND DEMIURGE

AI AND THE LEGAL STATUS OF AI CREATIONS

Camille Alegre¹

1. AN INHERENT CONTRADICTION BETWEEN CREATIVITY AND AUTOMATION

The recent DABUS case has shed light on a fundamental disagreement over how the law should understand inventorship. Whereas Australia² and South Africa³ have ruled that an artificial intelligence (AI) may qualify as an inventor, the United Kingdom,⁴ Canada,⁵ the United States,⁶ and the European Patent Office⁷ have ruled otherwise. Although the problem cannot be framed as clearly for copyright (which traditionally does not require the same amount of paperwork as a patent), similar questions and disagreements have been raised by decisions to grant co-authorship to an AI application in India (Sakar, 2021). To explore various legal ways to protect AI creations, there have been consultations by the WIPO (*WIPO Conversation on Intellectual Property and AI of 2020*) and in several countries,

-
1. Juriste IP/IT/GDPR chez DSM Avocats à la Cour au Luxembourg.
 2. *Thaler v. Commissioner of Patents* [2021] FCA 879.
 3. *Companies and Intellectual Property Commission : Patent Journal Including Trademarks, Designs and Copyright in Cinematograph Films*, vol. 54, n° 07, dated July 2021.
 4. *Thaler v. Comptroller General of Patents Trade Marks and Designs* [2021] EWCA Civ 1374.
 5. The Canadian Intellectual Property Office has not issued a final decision yet, but in a letter dated November 8, 2021 it asked the applicant to identify himself as the legal representative of the machine (Penner, 2022).
 6. *Decision on Petition regarding application* n° 16/524,350.
 7. EPO decision of 27 January 2020 on EP 18 275 163 and EPO decision of 27 January 2020 on EP 18 275 174.

including Canada (*A Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things*, 2021).

But how is AI challenging the current legal framework of intellectual property? And how have the UK and Australia, two common law countries, reached totally opposite conclusions despite identical premises?

Before investigating the origins of legal disagreement, and the reasons, we need to clarify some of the key concepts used throughout this chapter. In particular, we should understand a sometimes overlooked aspect of this issue: a lack of standardized terminology.

1.1 Key Definitions

We will define two concepts: intellectual property; and artificial intelligence.

1.1.1 Defining Intellectual Property

Intellectual property (IP) is a relatively recent umbrella term that encompasses different rights, such as copyright, trademark, and patent laws. Stallmann says it originated with the founding in 1967 of the World Intellectual Property Organization (Stallmann, 2006). This use of the word “property” draws a parallel between property law and intellectual property rights. Intellectual property is therefore a defined and finite resource that is appropriated by an entity (the owner). It is “intellectual” because it is created through mental activity.

In other words, the concept has two core features: (1) a finite resource (the creation); and (2) a mental activity that leads to production of the resource.

Both features are key to understanding the intersection between intellectual property and AI, as discussed in this chapter.

1.1.2 Defining AI

What is AI? To date, there has been no standardized definition. Although lack of standardization is common in science (Wang, 2019), it is particularly problematic in law and may cause confusion and false conceptions (Krafft, 2019).

Boucher described AI as an “umbrella concept” that encompasses various technologies (Boucher, 2020). Coined in 1956 by John McCarthy

and Marvin Minsky, the term was designed to reflect a new development in cybernetics: the rise of symbolism (i.e., an approach based on human intelligence and its mode of operation) in opposition to the hitherto dominant “connectionism” (Cardon, 2018). With this term, McCarthy and Minsky were referring to cybernetics not only in its current state but also in its future potential to mimic the human mind (Leufer, 2019). Because the term was coined to describe an ambition for the future, rather than a current phenomenon, it has ended up creating an ontological uncertainty. The current definitional problem arises from the two core features of symbolism: (1) the parallels it makes with the human mind (Scherer, 2016); and (2) the confusion over whether it refers to a current state or a goal. These two defining features have structured current debate on artificial intelligence.

1.2 Incompatibility between AI and IP

The fundamental incompatibility between AI and IP is not legal but terminological. We will pursue this argument by assessing the compatibility of the above two features with each other and by studying the metaphors and images associated with each of them, in line with the work of Lakoff and Johnsen (Lakoff, 1980).

The next section will argue that the anthropomorphic definition of AI prevents us from seeing the incongruity between AI and true mental activity. This incongruity is the main reason why authorship should not be extended to AI. The third section will turn to the compatibility of property with resources, and the time uncertainty in AI. Progress is interlinked with creation through a shared perception of time and innovation.

Finally, the last section will analyze the intersection between mental activity and time uncertainty, concluding that the issue of authorship may be the new Turing Test for intelligent machines.

1.3 Final Preliminary Remarks

“Author” and “authorship” will be used throughout this chapter as generic terms that encompass both “inventor” under patent law and “author” under copyright law (both will be defined in the relevant sections).

Although some of the analyses could be extended to other jurisdictions, the focus will be on European, American, and Canadian legal systems. The issue of trademarks will not be covered, as this right has so

far been left out of legal debate on AI and does not seem to raise specific difficulties (on this, see, for instance, the WIPO conversations mentioned above or Hartmann, 2020.)

2. ANTHROPOMORPHIC BIAS AND ARGUMENTS AGAINST THE LEGAL EXISTENCE OF AI CREATORS

Intelligence and intellectual work are both considered to be specifically human (sections 2.1 and 2.2.). In this debate, the criterion of mental activity is thus key to distinguishing between creations that can be protected and those that cannot.

2.1 Intelligence as a Specifically Human Feature

An anthropomorphic vision of AI is deeply embedded in the term itself because “intelligence” is commonly considered to be specifically human (Wang, 2019). This term lacks, however, a standardized definition within the scientific community (Scherer, 2016).

Identifying an entity as intelligent is not so much a technical impossibility as a semantic one. If intelligence is specifically human, “artificial intelligence” is, at best, an oxymoron and nothing more (Julia, 2019).

Demonstrating creativity may therefore be the new Turing Test for proving the intelligence of a machine. But “creativity” is no less semantically problematic than “intelligence,” as will now be shown.

2.2 Originality and Inventiveness in Intellectual Property

Mental activity is key to recognition of authorship, as shown by two main concepts of law on intellectual property: “originality” in copyright law; and “inventiveness” or “non-obviousness” in patent law.

2.2.1 Originality as a Specifically Human Feature

Although the different conventions on intellectual property rights seldom explicitly state that an author should be a natural person, traces of this assumption can be found in different legal systems and ultimately originate in the Platonic perception of art and creativity.

2.2.1.1 *Defining Originality from a Legal Standpoint*

In the European Union, case law has stressed the importance of originality in the final creation, which must result from “free and creative choices.” This concept is also found in Canadian case law, as the Supreme Court of Canada has ruled that creation cannot be the result of a “purely mechanical exercise,”⁸ and in American law, under which randomness alone cannot be considered to be original.⁹ This opposition between free choice and mechanical exercise seems to imply that creativity can be only a human characteristic. This inference is supported by the 1886 *Berne Convention for the Protection of Literary and Artistic Works*, particularly the reference in Article 7 to the author’s death and the reference in Article 6 to the author’s honour and reputation. In a report published in 2020, the United States Patent and Trademark Office made its position clear: AI cannot be granted authorship under US copyright law.¹⁰

2.2.1.2 *Originality in the Philosophy of Art*

In philosophy, the “original” is often opposed to the “copy.” This duality can be traced back to Plato. In *The Republic*, he draws a fundamental distinction between the “original” (the first and genuine creation) and the “mimesis” (the “fake” or copy). The second concept is reflected in common conceptions of AI, which is perceived as a device that mimics human mental processes (Wang, 2019). This perception is even a pillar of the symbolic AI approach and influences our current vision of AI (Cardon, 2018). If the creative processes of AI are only a matter of imitation, how could they result in an original work?

2.2.2 *Inventiveness as Another Specifically Human Feature*

Although the concept of “inventor” is not always defined in patent law, there is now strong evidence that European, American, and Canadian legal systems support the idea that an inventor should be a natural person. Moreover, the mental activity requirement, usually included in the criterion of inventive step or non-obviousness, has to be measured in terms of human skills. Assessing AI processes in terms of human ones can only make that criterion ontologically unreachable.

8. *CCH Canadian Ltd. v. Law Society of Upper Canada*, [2004] 1 S.C.R. 339, 2004 SCC 13.

9. *Naruto v. Slater*, n° 16-15469, 9th Cir. (2018).

10. Public Views on Artificial Intelligence and Intellectual Property Policy.

2.2.2.1 *The Inventor as a Natural Person*

US patent law defines an inventor as an “individual” and therefore *de facto* excludes AI from the definition of “inventor” (35 U.S. code § 100 f.). Moreover, the United States Court of Customs and Patent Appeals has stated that an invention involves distinct stages, one of which is the “mental part of the inventive act” (*Townsend v. Smith* 36F 2d 292 (C.C.P.A. 1929)). The adjective “mental” necessarily refers to a human being by referring to the activity of a mind.

Other patent laws may be less explicit. Article 60 of the 1975 European Patent Convention and Section 27(4) of the 1985 Canadian Patent Act both mention the “inventor” without defining the term. Some clarification has been provided by the recent rulings of the Dabus cases : for the above two legal texts, an inventor should be a natural person.¹¹

2.2.2.2 *Measuring AI Inventiveness in Terms of Human Creativity*

“Intellectual work” implies comparison with human beings because it is commonly understood as a specifically human feature. Therefore, AI creativity almost inevitably has to be measured in terms of human creativity. This leads to an aporia : if the threshold is a human characteristic, AI can never meet it, as this characteristic would therefore cease to be specifically human. This is a problem only with the symbolic AI approach, which holds that AI mimics human processes. It is not a problem with the connectionist AI approach (Cardon, 2018).

In patent law, it is difficult to assess AI creativity in human terms because the assessment has to be performed by a person who is skilled in analyzing the non-obviousness of an invention (see for instance Art. 35 U.S.C. 103 of the US Patent Code or Section 28.3. of the Canadian Patent Act). To that end (in line with Article 56 of the European Patent Convention), the European Patent Office has to verify whether the invention offers a “solution to a technical problem” (Chapter 3, Section 4 G-VII of the 2021 European Patent Guide 21st edition). As the creative processes of AI are unclear, it may be sometimes difficult to identify the problem tackled by the invention (the metaphor of the black box is one of the most commonly used with reference to AI [Foryciarz, 2020]).

11. See footnotes 5,6,7.

2.2.3 *The Degree of Autonomy of Intelligent Machines*

It is important to distinguish here between AI-assisted creation and AI-generated creation, a distinction that appears in the *Canadian Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things*, and the *WIPO Conversation on Intellectual Property and AI*. From a purely technical standpoint, no distinction is possible, as machine autonomy cannot be understood in the same way as human autonomy (Johnson, 2017). A distinction can be made, however, from the standpoint of intellectual property. An AI-assisted creation can be protected (the criteria of mental activity is applied to the human creator), but an AI-generated creation cannot (Johnson, 2017). This may for instance explain why Ankit Sahni's AI-assisted creation was accepted for registration in India, although the Indian registration system does not involve a preliminary examination of the protected work. The decision could therefore be legally challenged (Sarkar, 2021).

The human being behind the creation thus has to be identified, even if the relationship between the two is loose. This may explain why certain jurisdictions have simplified the requirement through legal fictions. The UK and Ireland, for instance, have created legal presumptions as to the authors and/or owners of computer-generated works. In Ireland, the owner is the person "arranging for creation of the work to be made" (*Irish Copyright and Related Rights Act*, 2000, Chapter 2, Section 21 (f)). In the UK, the owner is the "person who undertook the arrangements necessary for its creation" (*Copyright Design and Patent Act*, United Kingdom, 1988, Chapter 1, Section 9 (3)).

2.3 **Loopholes**

The above legal approach can be challenged in two ways. First, we could redefine the concept of originality less tautologically, thus creating room for protection of AI. Second, we could challenge the need to identify a human behind the creation, on the grounds that an AI creation is the result of work by many different stakeholders.

2.3.1 *Redefining Originality*

In the last decade of the 19th century, Marcel Proust, a French writer, wrote a text challenging the hegemonic vision of creativity in the literature of that time (Proust, 1954). To him, a creation should be always detached

from its creator's personality and understood as an entity in its own right. It does not reflect its author's views or past and should be understood as a finished product in and of itself. Originality should be sought not in the author's mental activity but rather in the audience's assessment of the creation.

Proust's argument implies that a creation could be protected even if there is no way to identify its author or fulfil the originality criterion. As will be discussed in the next section, such protection could be achieved through different legal instruments. These instruments, or similar ones, may be designed and used to protect AI-generated works in cases where identification of a human creator is impossible.

2.3.2 *Collective Creations*

The author/inventor's importance raises special problems in the case of artificial intelligence, which is usually a result of work by many different actors—from computer scientists to software architects and developers.

The multiplicity of actors creates overlapping rights to intellectual property. Different aspects of one invention may be protected by different rights. Also, the task of identifying a human creator may result in the same individual receiving intellectual property rights twice: once for inventing the AI application, and again for the creation generated by the application.

In a situation where a creation involves many stakeholders, it is also difficult to identify the actual creator. A multiplayer approach is sometimes advised, but may be unusable in practice when there are too many stakeholders (Shlomit, 2017). The next section will be about such overlapping rights and the difficulty in identifying all the stakeholders.

3. A CREATION WITHOUT A CREATOR

Intellectual property is also characterized by “ownership.” Using this concept, we can circumvent the difficult criterion of authorship and view creations as resources that can be appropriated. Whereas authorship is legally justified by philosophical arguments, ownership is legally justified by economic arguments that rely primarily on the concept of “competition.” However, the latter arguments seem to rely on a conception of AI rooted not in the present but in the future.

3.1 Creations as Resources in the Race to Develop AI

“Property” implies the existence of a resource that can be appropriated. In the case of AI, the resource can be either the data used by the application, or the AI itself and the creations it generates.

3.1.1 Data as the New Oil

The new hype in AI is due to the recent emergence of big data. Indeed, such technologies as machine learning and deep learning have been made possible by a huge increase in easily accessible data—now called the “new oil” (Hitch, 2014). In many cases, however, the data are already owned and legally protected. Such cases shed light on a concept common to intellectual property and AI: competition for a resource.

3.1.1.1 Ownership of Data

Although not all AI technologies require data, this is certainly the case with machine learning and deep learning—two technologies that have been booming for the last two decades. Both use data for supervised or unsupervised learning. In the first case, data ownership is clearly identified. In the second case, identification is more difficult.

Given the diverse uses of AI applications, the data can be of all types. Some data may already be protected by copyright laws or other laws on intellectual property. In this case, the use may infringe on existing property rights.

Training data may include personal data. Such use does not always comply with current privacy laws, particularly the principle of transparency. Indeed, this is one of the core principles of the 2016 *General Data Protection Regulation* in the European Union¹² and Schedule 1 of the 2000 *Canadian Personal Information Protection and Electronic Document Act*. Under this principle, data subjects must be informed of the intended use of their personal data, although such use, and its impact on the final product, is unclear when an AI is being developed.

To circumvent the risks of noncompliance, some have suggested extending the existing exceptions to the above two regulations. One exception is

12. EU, *Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*, [2016] JO, L 119/1.

fair use (Patel, 2020), and the other is research purposes (Ivanova, 2019). The assumption is that privacy and intellectual property should not be absolute and may be overridden for the development of AI. In other words, progress should have priority over individual rights. This stance arises from a perception of technology that is discussed in Section 3.2.

3.1.1.2 *Appropriating Data for AI Applications*

In some countries, legal protection exists not only for data but also for the actions of collecting and organizing data. The European Union has put forward an original, and *sui generis* legal protection for databases (*Directive 96/9/EC on the legal protection of databases*). It covers the person who compiles the databases, regardless of whether the act of compiling has a creative dimension (Article 3 of the directive). It is still unclear how the directive applies to machine-generated data (Bond, 2021). This legal protection deserves to be mentioned insofar as it shows that lawmakers have sometimes freed patent and copyright regulations from the mental activity requirements.

3.1.2 *Ownership versus Authorship*

The “new oil” metaphor depicts data as resources that may be appropriated. To encourage innovation, data should be allowed to flow freely, as stated by the European Union (Communication from the European Commission, *A European Strategy for Data* COM/2020/66 final.). This broader view of the benefits of data creation is not limited to AI training data and could be extended to the outputs of AI. Indeed, the language of innovation in general, and of AI in particular, revolves around the idea of “conquest.” Granting authorship to AI is thus no longer a way to acknowledge AI creativity but also a way to encourage investment in innovative technologies.

3.1.2.1 *Conquering Innovation*

Launched in 2014, two AI startups, Qentis and Cloem, have shown a new way to consider IP. Qentis aims to conceive almost every possible text of 400 words, and Cloem seeks, for example, to generate as many nucleotide sequences as possible (see for instance George, 2015). This obsession for quantity can be understood only within the context of economic competition. Both Qentis and Cloem are seeking to invent something before

someone else does. The underlying assumption is that an invention is a finite resource that users compete for.

This view of an invention also appears in the metaphors of (1) property and (2) a race competition. If the acquisition of property is, by definition, the appropriation of a finite resource, creations should be viewed as resources as well, since they are referred to as “intellectual” property (Stallmann, 2006). IP rights most visibly have economic consequences in the case of patenting because a patent usually confers a monopoly. Even copyright has economic consequences through licensing and the payment of royalties.

Using the metaphor of AI as a race competition, Cave and ÓhÉigeartaigh have pointed to the danger of prioritizing speed over ethics and governance (Cave, 2018). Mapadimeng has also highlighted the deterministic vision of progress that runs through discourse on the Fourth Industrial Revolution (Mapadimeng, 2019). The creation of an object is thus seen as inevitable, and the creator is no longer the one who shows originality but rather the one who is fastest in the creative process. As a consequence, the author’s identity does not matter. What matters is support for the creative process.

3.1.2.2 *Encouraging Investment through IP Rights*

Granting authorship to AI encourages innovation by providing an economic incentive. This is the view of Abbott, a key scholar in the Dabus case. Indeed, Abbott has argued that AI authorship would incentivize development of creative computers (Abbott, 2016).

Many current IP systems already allow for the possibility of distinguishing between ownership and authorship (e.g., when an employee makes an invention at work)¹³ as well as the possibility of erasing authorship (e.g., to protect anonymous artistic work).¹⁴ Relevant to this discussion is the economic context within which Australia and New Zealand have granted AI the status of “inventor.” The AI market is already dominated by a few powerful players (Ulnicane, 2020). Is it really desirable to restrict citizens’ privacy and other rights so that a few companies may gain an even greater market advantage?

13. See for instance Article 60 of the *European Patent Convention*.

14. *Berne Convention for the Protection of Literary and Artistic Works*, 828 RTNU 221 September 9, 1886 [amended September 28, 1979], Art. 7.3 and 15.

3.2 Innovation as an Unstoppable Force

Innovation through IP is perceived as being economically unstoppable. Because IP rights allow transparency, they are often claimed to be a lesser evil than trade secrets as a means to protect the inventions that will inevitably arise. By believing so optimistically in progress, however, we may be mistaking the imagined future of science fiction for the actual present.

3.2.1 Trade Secrets as a Deterrent

For Morozov, the values embodied in the Internet have slowly spread through the whole of society (Morozov, 2013). In particular, values of openness have emerged from computer science with the creation of new legal instruments to promote transparency, such as the open-source movement and creative commons licensing. Whether these new instruments should be preferred to traditional IP rights is beyond the scope of this chapter. In any case, they illustrate that, in the knowledge economy, sharing requires transparency. Trade secrets are thus perceived as a threat to the free flow of knowledge. This view should nonetheless be nuanced: IP rights are far from a guarantee of transparency, and some of the perceived opacity of AI arises from copyright protection (e.g., Castelluccia, 2019).

3.2.2 An Uncertain Future

As mentioned in the introduction, the term AI carries an ambiguity. Does it refer to the present or the future? This confusion is sometimes deliberately cultivated by certain specialists, who fear a new AI winter, and by the portrayal of AI in the media (Marcus, 2019). It can even be observed in the *Canadian Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things*, whose definition of AI encompasses general AI, a concept that, to date, belongs only to science fiction (e.g., Brooks, 2017). Such exaggeration could lead to adoption of an unsuitable legal framework and, even more concerning, to disregard for current problems with current uses of AI (Boucher, 2020).

The same ambiguity is present in the language of the Fourth Industrial Revolution (Mapadimeng, 2019). Mapadimeng has shown that commentators view themselves not as contemporaries of this change but rather as observers in a hypothetical future. This viewpoint may lead lawmakers to disregard certain societal impacts of AI innovations. Choosing innovation as the highest value is a political choice that should be made democratically,

and not dictated out of a sense of emergency and helplessness in the face of an unstoppable technology.

4. THE (SCIENCE) FICTION OF CREATIVE AI

Despite interesting developments over the past few years, it seems that granting AI the status of creator will not (and perhaps should not) happen overnight in most countries. In addition to the reasons given in Section 1, reluctance to grant such status may also arise from biased social representations. But why does this issue still attract so much attention? The answer may be its impossibility. An AI is not a creator, despite efforts by computer scientists to make it so.

4.1 AI as a Bogeyman : Social Reluctance and Perception of AI

Cultural representations, often flawed, are having an impact on social perceptions of AI (e.g., The Royal Society, 2018; George, 2015). For instance, AI-related articles show a significant mismatch between their content and their illustrations (Geitgey, 2018). An ambient AI anxiety is also being fuelled by certain recurring myths, such as the technological singularity (Brooks, 2017; Johnson, 2017). Granting authorship to AI may thus have not only economic significance but also a mainly symbolic one. It would be perceived as an acknowledgment that AI can do something previously thought to be purely human.

4.2 Creativity as the Ultimate Turing Test for AI

We have provided evidence for the semantic impossibility of viewing AI as creative. But the impossibility may be more than definitional: development of a technology is not straightforward and often has unexpected consequences (Debray, 2006). The issue itself may have value only as long as the standard is unreachable.

4.2.1 *Paradox of the Last Mile of Automation*

From the beginning, AI scientists have had a propensity to announce ground-breaking innovations that never occur (Cardon, 2018). This propensity is even one of the main reasons for the different AI winters that have punctuated the history of AI. Automation has shown a similar trend: as more and more processes become automated, there have arisen more and more automation-related human tasks: maintenance, support,

etc. (Gray, 2015). A similar trend may be occurring in the automation of creativity. Indeed, as AI applications develop, the creative process has expanded to encompass more and more human stakeholders (including digital labour). A creative process emancipated from all human work may therefore be nothing more than a fiction.

4.2.2 *AI as Indicative of the Creative Process*

Different tests have been designed to measure the intelligence of a machine. The most famous one is indubitably the Turing Test, but there are some other, more practical tests, such as the ability to play Go or chess. In particular, by using chess to assess AI intelligence, researchers have learned about the complex cognitive processes it requires (Ensmerger, 2011). By using authorship as the ultimate and unreachable goal for AI, we could thus better understand human creativity. To end the present chapter, we may conclude that (1) creation is not an individual process; and that (2) concerns about AI may reflect the anxieties of our time more than the real threat of a technological singularity.

Why should creation be considered a collective process? Because it implies a constant dialogue between past creations and present ones. AI achieves such dialogue through the use of training data. Human writers achieve it through intertextuality (Kristeva, 1980).

In Section 3, we have shown that a deterministic approach to innovation may imply that creations are inevitable. But the theory that ideas have a life of their own is not new and has notably been advanced by Foucault (Foucault, 1969).

REFERENCES

Legislation

Treaty

Berne Convention for the Protection of Literary and Artistic Works, 828 RTNU 221 September 9, 1886 [amended September 28, 1979].

European Legislation

European Patent Convention, 1973.

EU, *Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*, [2016] JO, L 119/1.

National Legislation

Australia

Thaler v. Commissioner of Patents [2021] FCA 879.

South Africa

Companies and Intellectual Property Commission : Patent Journal Including Trademarks, Designs and Copyright in Cinematograph Films, vol. 54, n° 07, dated July 2021.

United Kingdom

Thaler v. Comptroller General of Patents Trade Marks and Designs [2021] EWCA Civ 1374.

United States

Decision on Petition regarding application n° 16/524,350.

Jurisprudence

Canada

CCH Canadian Ltd. v. Law Society of Upper Canada, [2004] 1 S.C.R. 339, 2004 SCC 13.

United States

Naruto v. Slater, n° 16-15469, 9th Cir. (2018).

Townsend v. Smith 36F 2d 292 (C.C.P.A. 1929).

Doctrinal references

Abbott, Ryan, "I Think, Therefore I Invent: Creative Computers and the Future of Patent Law", *Boston College Law Review*, vol. 57, n° 4, 2016, p. 1079-1126.

Bond, Toby, "Commission Inception Impact Assessment on a Proposed Data Act: What Does it Mean for IP Owners?", *Bird & Bird*, 2021, <<https://www.twobirds.com/en/news/articles/2021/uk/commission-inception-impact-assessment-on-a-proposed-data-act>>.

Boucher, Philip, "Artificial Intelligence: How Does it Work, Why Does it Matter, and What Can We Do About it?", *European Parliamentary Research Service*, n° 56, 2020.

Brooks, Rodney, "The Seven Deadly Sins of Predicting the Future of AI", *MIT Technology Review*, November 2017.

Cardon, Dominique, Jean-Philippe Cointet and Antoine Mazieres, « La revanche des neurones : L'invention des machines inductive et la controverse de l'intelligence artificielle », *La Découverte Réseaux*, vol. 5, n° 211, 2018.

Castelluccia, Claude and Daniel Le Métayer, "Understanding Algorithmic Decision-making: Opportunities and Challenges", *Panel for the Future of Science and Technology*, 2019.

Cave, Stephen and Seán S. ÓhÉigartaigh, "An AI Race for Strategic Advantage: Rhetoric and Risks", *AAAI / ACM Conference on Artificial Intelligence, Ethics And Society*, 2018.

Debray, Régis, « L'effet Jogging », *Conférence à Seville*, 2006.

Ensmenger, Nathan, "Is Chess the Drosophila of Artificial Intelligence? A Social History of an Algorithm", *Social Studies of Science*, vol. 42, n° 1, 2011.

Foryciarz Agata, Daniel Leufer and Katarzyna Szymielewicz, "Black-Boxed Politics: Opacity is a Choice in AI Systems' Medium", *Medium*, 17 January 2020, <<https://medium.com/@szymielewicz/blackboxed-politics-cebc0d5a54ad>> (accessed 15 July 2020).

- Foucault, Michel, *L'archéologie du savoir*, Paris, Gallimard, 2008 [1969].
- Geitgey, Adam, "The Real Scandal of AI: Awful Stock Photos", *Medium*, 2018, <<https://medium.com/@ageitgey/the-real-scandal-of-ai-awful-stock-photos-456633b9b0fc>>.
- George, Kiran, "The Curious Case of Qentis and Cloem", *Spicy IP*, 2015, <<https://spicyip.com/2015/02/the-curious-case-of-qentis-and-cloem.html>>.
- Gray, Mary L., "The Paradox of Automation's Last Mile", *Social Media Collective*, 2015.
- Hartmann Christian, Jacqueline Allan, P. Bernt Hugenholtz *et al.*, *Trends and Developments in Artificial Intelligence Challenges to the Intellectual Property Rights Framework*, Luxembourg, Publications Office of the European Union, 2020.
- Hitch, Dennis, "The Glass House Effect: Big Data, the New Oil and the Power of Analogy", *Maine Law Review*, vol. 66, n° 2, 2014, p. 373-395.
- Ivanova, Yordanka, "Re-using Personal Data for Statistical and Research Purposes in the Context of Big Data and Artificial Intelligence", *Annual Privacy Forum 2019: Privacy Technologies and Policy*, 2019.
- Johnson, Deborah G. and Mario Verdicchio, "Reframing AI Discourse", *Minds and Machines*, vol. 27, n° 4, 2017, p. 575–590.
- Johnson, Deborah G. and Mario Verdicchio, "AI Anxiety", *Journal of the Association for Information Science and Technology*, vol. 68, n° 9, 2017.
- Julia Luc, *L'intelligence artificielle n'existe pas*, Paris, First edition, 2019.
- Krafft, P.M., Meg Young, Michael Katell, Karen Huang and Ghislain Bugingo, "Defining AI in Policy versus Practice", *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 2019.
- Kristeva, Julia, *Desire in Language: A Semiotic Approach to Literature and Art*, New York, Columbia University Press, 1980.
- Lakoff, George and Mark Johnsen, *Metaphors We Live By*, Chicago, University of Chicago Press, 1980.
- Leufer, Daniel, "Dismantling AI Myths and Hype", *Mozilla Fellowships & Awards Blog*, 2019.
- Mapadimeng, Mokong S., "The Fourth Industrial Revolution (4th IR): A Theoretical Critique of the Discourse", Paper presented at the *South African Sociological Association Annual Congress*, Southern Sun Hotel, Pretoria, 15-17 July 2019.
- Marcus, Gary, *An Epidemic of AI Misinformation*, The Gradient, 2019.
- Morozov, Evgeny, *To Save Everything, Click Here: The Folly of Technological Solutionism*, Public Affairs New York, 2013.
- Patel, Shaleen J. and Sushmitha Rajeevan, "Artificial Intelligence and Copyright: Ownership and Fair Use", *IP Strategist*, June 2020.
- Penner, Marc D., "Canada: AI not an "Inventor" under European Patent Law: Is Canada Heading Down the Same Path?", *Mondaq*, 2022, <https://www.mondaq.com/canada/trademark/1159636/artificial-intelligence-not-an-inventor-under-european-patent-law-is-canada-heading-down-the-same-path#_ftn2>.
- Proust, Marcel, *Contre Sainte-Beuve*, Paris, Gallimard, 1987 [1954].
- Sarkar, Sukanya, "Exclusive: India Recognises AI as Co-author of Copyrighted Artwork", *Managing IP*, 2021, <<https://www.managingip.com/article/b1t0hfhz2bytx44/exclusive-india-recognises-ai-as-co-author-of-copyrighted-artwork>>.

- Scherer, Matthew U., "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies", *Harvard Journal of Law & Technology*, vol. 29, n° 2, 2016.
- Stallmann, Richard, "Did You Say 'Intellectual Property'? It's a Seductive Mirage", *Policy Futures in Education*, vol. 4, n° 4, 2006, p. 334-336.
- The Royal Society, *Portrayal and Perceptions of AI and Why They Matter*, London, Royal Society Publications, 2018.
- Ulnicane, Inga, William Knight, Tonii Leach, Bernd Carsten Stahl and Winter-Gladys Wanjiku, "Framing Governance for a Contested Emerging Technology: Insights from AI Policy", *Policy and Society*, vol. 40, n° 2, 2021, p. 158-177.
- United States Patent and Trademark Office, *Public Views on Artificial Intelligence and Intellectual Property Policy*, 2020.
- Wang, Pei, "On Defining Artificial Intelligence", *Journal of Artificial General Intelligence*, vol. 10, n° 2, 2019, p. 1-37.
- Yanisky-Ravid, Shlomit and Liu Xiaoqiong, "When Artificial Intelligence Systems Produce Inventions: An Alternative Model for Patent Law at the 3A Era", *Cardozo Law Review*, vol. 6, n° 2215, 2017.

CREATIVE PIRACY AND AI POLICING OF INTELLECTUAL PROPERTY

Stephen Monteiro¹

Digital media users have been duplicating and sharing copyrighted programs and files without authorization for decades. In a defining moment from the dawn of personal computing, Bill Gates published an open letter in 1976 to computer hobbyists, admonishing them for making and sharing unauthorized copies of BASIC, Microsoft's first programming product. "Most of you steal your software. [...] Is this fair?" Gates asked. To curb unauthorized use, Microsoft would devise multiple licensing schemes and anti-piracy measures in the following decades, including disk holograms and product keys. With the development and popularization of the World Wide Web in the 1990s, however, sharing of digitized media files—from software to music albums—became a common, easy transaction among networked computer users on a global scale. As mainstream media production and consumption has shifted to online platforms with smart devices and comprehensive streaming services, unauthorized file-sharing—particularly of commercial audio and visual material, such as songs and movies—has become an intractable problem for media companies.

In recent years, intellectual property management firms and internet media-sharing platforms, like YouTube and Facebook, have turned to artificial intelligence to become more effective at detecting and combatting unauthorized distribution and use of copyrighted audio and video works.

1. Assistant Professor of Communication Studies at Concordia University.

Complementing other technologies, such as digital rights management access tools, this kind of automated system relies on machine learning to identify specific traits of copyrighted works and the files that bear them. It may also assess their online context and exposure, and determine their source and intent of use. In response to such systems, individuals who reproduce and share copyrighted works without authorization sometimes resort to masking techniques to distort or disguise content in ways that may mislead AI while preserving enough of the original sound and image quality to remain appealing to human audiences. This practice could be termed “creative piracy.” It involves (sometimes major) aesthetic alterations to the legible symbols and codes of original works in ways that could be considered transformative, thereby creating new works potentially protected by fair use provisions of copyright law. By developing specific audiovisual alteration techniques to evade machine detection, such online pirates raise important questions about fair use in the face of automated decision-making and about what constitutes an illegal copy of a work. If an altered work could fool the eyes and ears of anti-piracy AI, for example, might it be argued that the changes are sufficiently transformative to constitute fair use? These techniques may also amount to a new form of remix, representing yet another way that networked users appropriate and alter existing cultural works, and thus producing new forms of online consumption while expanding digital media aesthetics. These creative alterations to copyrighted works can be related to the practices involved in making memes, mashups, and other distinctly digital forms of appropriation art that have become an essential component of everyday online interactivity and communication (Sonvilla-Weiss, 2010).

We will explore the relationship between AI copyright management and anti-piracy AI through the lens of broader issues of digital aesthetics, vernacular cultural production, and fair use. Anti-piracy AI, and the assumptions that govern its operations, may inadvertently lay the groundwork for the creation of derivative (pirated) works that challenge creative—and potentially legal—norms surrounding intellectual property. By using numerous filters, effects, and editing tactics, the producers of these pirated works may sufficiently distort and degrade the original material as to produce a different work. While there have been vigorous debates about the possibility of copyrighting AI-produced artistic works (Ligon, 2019; Berton, 2020; Sobel, 2017; Vézina and Moran, 2020), little attention has been paid to the alteration and potential transformation of existing copyrighted property in response to the automated “audiences”

of anti-piracy AI. This chapter will first present a brief historical trajectory of online file-sharing's relationship to copyright and digital creativity. That presentation will be followed by a discussion of automated content moderation via AI systems—especially YouTube's Content ID—that detect and curb unauthorized posting of copyrighted materials on social media and sharing platforms. Attention will then turn to the creative tactics used by uploaders to alter the aesthetic qualities of copyrighted audiovisual works, with analysis of specific examples of altered TV and film works posted to YouTube. These examples will demonstrate the difficult balancing act of attempting to make works unrecognizable to AI systems but still acceptable to viewers. We will conclude with an assessment of how this relationship between AI and creative piracy may impact fair use and the aesthetics of media appropriation and alteration.

1. DEVELOPMENT OF ONLINE PIRACY AND CREATIVE PRODUCTION

The 1990s saw the development of decentralized digital user networks through the popularization of the internet as well as increased digitization of written, musical, and audiovisual works. These two trends together led to new forms of distributing and accessing cultural works at the turn of the 21st century. This set the stage for extensive, routine reproduction and unauthorized sharing of copyrighted works on a global scale. Although unauthorized copying of cultural works existed for centuries (Johns, 2011), it would become easier and more common during the 20th century as cultural production became increasingly mechanized for mass production. By the 1980s, books, music, TV programs, and films were commonly copied and shared through analog duplication processes. People would record music on audio cassettes, and TV programs or films on videocassettes, while duplicating books and other publications via photocopiers. Although these activities were often in violation of copyright laws against unauthorized reproduction, they were generally tolerated by media industries when practised by private individuals. Such tolerance existed for at least two reasons. First, the material conditions of the recording technologies (machines loaded with tapes or paper) made large-scale duplication cumbersome and time-consuming. Second, such copying generally produced lower-quality reproductions (i.e., some noise and distortion inevitably arose during the copying process), which were less likely to impact the market for high-quality, licensed works. Unauthorized copying of protected media during that period—outside large-scale piracy

operations that thrived in certain parts of the world—occurred primarily within small groups of friends. Teenagers would exchange audio cassettes at school, and adults would share films recorded from cable TV to play on their VCRs. This unauthorized copying sometimes included creative transformation of the original works. One example would be the music mix tape, a compilation of songs from different artists and albums often made by one person as a gift to another. These cassettes—a key part of North American youth culture in the late 20th century—would often be embellished with original recordings (e.g., the maker's voice) and handmade artwork for the cassette cover (Moore, 2005). Another example would be copy art, which contributed to fanzine and poster culture by producing original graphic works on photocopiers, often through the reproduction and rearrangement of excerpts from appropriated materials (Eichhorn, 2016).

Unauthorized copying and sharing was definitively transformed in the 1990s with the rise of digitized media products and the networking of computer users through public communication systems (such as telephone lines). As media companies distributed digital copies of their works—movies on digital video discs (DVDs) and music on compact discs (CDs)—these files could be downloaded to computer hard drives with little-to-no loss of quality. In turn, they could be shared with other internet users through peer-to-peer (P2P) websites, such as Napster or Kazaa, and by using specific communication protocols, such as Gnutella or BitTorrent, in conjunction with a protocol client on the computer. As governments, law enforcement, and media corporations attempted to dismantle these networks (such as the US court-ordered shutdown of Napster in 2001), unauthorized distribution of copyrighted works shifted to other platforms and practices. Nevertheless, the advent of new distribution models by media companies, including app-based online stores (e.g., iTunes) and streaming services (e.g., Netflix, Spotify), has helped curb file-sharing of mainstream products by offering low-cost, reliable, and secure alternatives to file-sharing.

As with the mix tapes and copy art of the late 20th century, the accessibility of the internet, the ease of file-sharing, and the availability of user-friendly editing software in the 21st century have brought new forms of user-based “remix” creative production, such as memes and mashups, that make use of the specific properties of digitally based works, electronic devices, and software (Navas, 2012; Jenkins *et al*, 2013). In addition, the copyleft movement and Creative Commons licences have introduced new, more flexible models for authorizing the reproduction and circulation

of original works protected by copyright laws (Berry, 2008; Harold, 2007; Lessig, 2008).

With the onset of social media as the main category of digital engagement and communication since 2005, new opportunities and challenges have arisen in the unauthorized circulation of copyrighted goods. By design, social media platforms strongly encourage individuals to share digital materials, especially images and videos, whether the materials are oriented primarily toward interpersonal interactions, such as Facebook, or the housing of content for mass sharing, as with YouTube. To avoid legal challenges that could lead to significant financial penalties or shutdowns, mainstream social media platforms encourage users to share content, while discouraging sharing of unauthorized material not only by including penalties in user agreements (e.g., banning repeat offenders from the platform) but also by resorting to pro-active, automated measures to police material shared on their sites.

2. AI CONTENT RECOGNITION AND PIRACY

Through the development of multiple technologies, cybersecurity and content recognition companies, such as Corsearch, Audible Magic, Red Points, Nagra-Kudelski, and Viaccess-Orca, are now offering automated anti-piracy services to rights holders. These companies frequently emphasize that AI outperforms human-based approaches in finding and combatting unauthorized availability and use of copyrighted materials. For example, Red Points claims to provide “artificial intelligence technology that automates the full process from detection to removal [for] scalable, limitless infringement takedown” (Red Points, 2023). There are usually two basic services. The first one is intense online monitoring that involves constant, automated crawling of the web in search of unauthorized storage, availability, or exchange of targeted copyrighted materials. In essence, bots scour online resources for distinctive data or metadata identifiers. Relevant data might be the code sequences that produce the particular images or sounds that make up the work. Relevant metadata might be code attached to files that do not constitute the work as such, but signal the origins or conditions of production or archiving of files. The second basic service is watermarking or tattooing. This involves embedding unique metadata code sequences into authorized copies of the work, which anti-piracy software can later seek out when searching for unauthorized copies. A forensic watermark or tattoo on an unauthorized copy helps investigators

not only identify the work but also trace the origin of the file with greater precision and potentially uncover security breaches. It is important to note that altering or degrading the displayed content of a file—for example, by laterally inverting a film scene or slowing down a soundtrack—will change the data of a file in addition to the appearance or sound, but it will not necessarily alter the metadata, including the forensic codes. Searches for metadata, such as watermarks, may thus uncover files of protected works even if the works have been greatly altered to evade recognition by automated systems.

Apart from these contractors, online piracy is also being fought by large platforms that allow or encourage users to upload and share material. These platforms have created their own AI-based copyright management systems to detect unauthorized use of materials. One example is Facebook's Rights Manager. Rights holders provide copies of their work to Facebook, which compares the copies with user-uploaded content. When matches are detected, the system will follow through with the rights holders' predetermined preferences to allow, block, or take down the matched content. YouTube's Content ID is a similar system and represents perhaps the best-known example of such anti-piracy AI designed for a single global platform. Content ID was added to YouTube in 2007 as Video Identification. It appeared only two years after the platform launch and had been developed to curb the regular uploading of copyrighted TV, film, and music content by YouTube users (King, 2007). As the internet's largest video-sharing platform, YouTube has over a billion users and about a billion hours of viewing per day with over 500 hours of new content uploaded every minute (Frangoul, 2018). Its simple uploading procedures—requiring only a free YouTube account—have made the platform popular for posting and sharing copyrighted materials, such as TV episodes, feature films, and music.

Before the existence of Content ID, if copyright holders wished to remove unauthorized materials, they had to notify YouTube on a case-by-case basis, a complicated and time-consuming process, especially for media companies with large catalogues of music, TV, or film holdings. These circumstances exposed YouTube to the risk of punitive lawsuits aimed at shutting down the platform for systematic copyright infringement under the *Digital Millennium Copyright Act* (DMCA). In 2001, the Recording Industry Association of America successfully sued Napster under the DMCA, thus ending that P2P platform's popularity (Gray, 2020, p. 83-88). As a proactive means to identify copyrighted materials, Content ID shifted

the burden of oversight from the rights holders and their agents to the platform itself.

Content ID requires media companies to submit reference copies of their protected works, against which the system compares the millions of daily video uploads to YouTube through fingerprinting tools. As with Facebook's Rights Manager, rather than automatically removing unauthorized uploads and sanctioning users, Content ID offers several options to rights holders. The latter can direct YouTube to remove all such material. They may also choose to permit uploads and circulation, in exchange for analytical data on users and uses, or monetize unauthorized use through YouTube's profit-sharing programs, which give rights holders a percentage of the advertising revenue generated through viewing of their property on YouTube (Edwards, 2018). "More often than not, content owners choose to earn revenue from the ads run against claimed videos," states YouTube, claiming that over \$3 billion in advertising revenue was passed along to rights holders during the first decade of Content ID's monetization option (YouTube Creators, 2010 ; Google, 2018).

3. NETWORKED CREATIVITY AND PIRACY

In the face of these efforts to use AI to recognize unauthorized availability and circulation of copyrighted video, audio, or audiovisual files, users hoping to share such materials—particularly on open platforms such as YouTube—have sometimes adopted different means of masking them, even in cases that would suggest fair use. "The only thing that matters is are you smarter than a robot [?]," stated one YouTube content creator in a 2020 report critical of Content ID's impact on fair use (Trendacosta, 2020, p. 5). Through creative piracy, i.e., by devising and integrating different visual and auditory effects into copies of copyrighted works, uploaders have attempted to throw off AI vigilance by simultaneously altering the aesthetic qualities of the audiovisual files while also trying to ensure that the changes are not so drastic as to curb the primary visual and auditory appeal, and thereby jeopardize audience interest.

In a 2018 report, YouTube claimed that its anti-piracy AI "can now catch efforts to evade detection like changing a video's aspect ratio, flipping images horizontally, and speeding up or slowing down the audio" (Google, 2018). Uploaders of unauthorized copyrighted materials have responded by further altering original works through heavy reediting of scenes and reframing of shots. Taken as a whole, these tactics represent

a fundamental change to the longstanding goals and practices of unauthorized reproduction. Historically, whenever individuals reproduced and circulated copyrighted works without authorization, they would seek to copy the original work as closely as possible, ideally making the copies indistinguishable from the original. Pirated books, for example, would include pirated dust jackets, and pirated audio cassettes would include pirated labels and cover art. That goal remains today in the case of fake luxury goods. For example, counterfeiters of Rolex watches and Chanel handbags try to make copies that fool even the expert eye.

Digitized works in digital networks can be reproduced almost instantly with virtually no loss of quality. Yet this attribute of digitally reproduced works also makes them easy to detect by automated systems trained to find and match them against the original files. Accordingly, to camouflage them effectively, online pirates resort to desktop editing and alteration techniques that have produced the flourishing remix culture of internet “prosumers”. Such practices, and the extent to which they alter existing works, raise important questions about creativity in digital culture and how online piracy may blur into production of new forms of audiovisual expression and meaning that could be protected in their own right or construed as derivative works.

Identifying derivative works, and the extent to which sampling of other works can be considered fair use, has become problematic in the digital era. With the rise of social media platforms, everyday social interaction and exchange through digital networks has involved the circulation and re-circulation of extant, generally copyright-protected digital material that may be produced by a large media corporation or an unknown amateur. Eduardo Navas, among others, has described such practices as “remix” (2012). Remix is a creative category that encompasses constant sampling, compositing, and reorganizing of single or multiple cultural works. In the early 21st century, it has contributed to the rise of Web 2.0 and participatory digital culture (Jenkins, 2006; Burgess and Green, 2009). It forms the foundation of the popular aesthetics of the internet, as social media users typically create and share a wide variety of amateur content by finding and altering material readily available online. Remixing existing content is an important aspect of fan culture, for example. Devotees of a celebrity, a work, or a media franchise will make and post tribute compilations (fanvids) to social media that may incorporate copyrighted film clips, music samples, and photographs (Freund, 2018). They will regularly use appropriated still and moving images to make or punctuate

social media posts, thereby converting all manner of copyrighted material (news clips, movie scenes, viral amateur videos) into JPEGs and GIFs that dot our social media feeds (Wagener, 2020).

While these participants in online participatory culture regularly traffic in copyrighted materials, they rarely create polished works, instead producing material of varying quality that can quickly strike the viewer as an amateur effort. Indeed, the poor quality of much reused material has become part of the aesthetics of social media production. Nick Douglas has called such production “internet ugly,” arguing that the apparent unskilled nature of such creativity contributes to a sense of its authenticity as the creation of an everyday, average media user. Douglas argues, “[r]ather than a genre, it’s a cultural dialect, used not just to frame certain propositional content but to communicate things about its user” (2014, p. 315).

Whereas such alteration of pre-existing—often commercially valuable—cultural works may be done for personal creative reasons, creative piracy seems to alter them strictly for the purpose of exploiting the potential limitations of automated surveillance. The deliberately degraded quality of copyrighted mainstream TV and film content shared on YouTube appears to participate in the “internet ugly” aesthetic identified by Douglas, but it does not seem to have the same propositional or communicative slant of other amateur works generated from copyrighted materials. The creative piracy meant to deceive AI monitoring recreates the work as if displayed in a funhouse mirror. Typical techniques include laterally inverting images, re-framing, shortening, or repeating shots, lightening or darkening tonality and contrast, slowing or accelerating playback speed, showing images from a shifted perspective (as though they were at an angle to the viewer), and even embedding images into other moving images or animated effects, such as flashes or waves of light. These changes test the limits of visibility and audibility, and thereby the limits of meaning-making itself. On the one hand, they transform the original material in fundamental ways by considerably altering the aesthetic choices of the original creators. On the other hand, they seek to leave enough of the narrative intact and legible so that users can access the intended meanings and effects of the original. The aim is twofold: a human user will be able to discern the meanings of the altered work despite various masking effects, but the automated surveillance systems will get caught up in the web of those added effects and thus miss the markers that identify the original material.

It is difficult to quantify the extent of these practices. YouTube does not provide statistics on the number of uploads or removals of unauthorized copyrighted works on its platform (Google, 2018). Nevertheless, if you search for films or TV series by using terms such as “full episode” or “full movie” or search for popular genres from the current year, you will typically find multiple videos that exploit such piracy techniques. These practices thus seem to be relatively common. Such videos may remain on the platform for weeks or even months before being removed, but they may also disappear after only a few hours or days. This impermanence makes it difficult not only to conduct sustained research on specific instances over an extensive period but also to cite examples in published research and expect them to remain available for the reader to consult independently. Nonetheless, some examples will be examined and described here as typical of the online piracy commonly encountered.

One example is a July 2021 upload of the 2013 film, *Homefront*. Using a different title and the description “Movie Powerful Action 2021 Full Length English latest New Best Action Movies,” it received 4.8 million views within three weeks, according to YouTube’s statistics, before being taken down by the uploader. YouTube offers the authorized version of *Homefront* for purchase or rental, and presumably the pirated video could have impacted the platform’s revenue from its authorized version. By simply viewing a short sample of this apparently unauthorized version, it was possible to understand the uploader’s principal masking technique. In each scene of the film, snippets of shots were briefly repeated or quickly intercut, seemingly jumbling diegetic space and time while preserving the basic continuity of the narrative. The original film is one hour and 40 minutes long, but this uploaded version was seven minutes shorter, indicating that shots or scenes had been removed to compensate for the repetition meant to fool Content ID (King of Special Forces, 2021). Similarly reedited was a copy of the Dwayne Johnson and Ryan Reynolds action comedy *Red Notice*, uploaded to YouTube in December 2021 and still available there in February 2022. One of Netflix’s most popular original film releases (Netflix), this version of *Red Notice* had over 156,000 views, but contained frequent ellipses of several seconds within every scene, thus shaving 37 minutes off the film length of one hour 58 minutes (Candance Celeste Stmarie, 2021).

Douglas has argued that, while “internet ugly” has meaning for social media users, the distortion and degradation tactics of online pirates may hamper human viewing experiences, yet likely have no impact on the

efficacy of large, dynamic anti-piracy AI systems, such as Content ID, to detect unauthorized use. “Maybe you suffered through it, recognizing that this degraded quality is a necessary sacrifice to avoid YouTube’s copyright bots. The bad news is, it was probably completely unnecessary,” says Douglas. “If content ‘gets past’ Content ID, it’s because the content owner isn’t looking for it—or has told YouTube to let infringing uploads stay up” (Douglas, 2018). In fact, in the case of the reedited *Homefront*, viewers often commented on the difficulties of viewing the film due to the masking technique. “Had to put my neck back in place from all the shifting I did watching this flick,” wrote one viewer. “I’m definitely going to have a headache for a few days,” wrote another. “Hatchet job of editing, bits and pieces, bits and pieces. [...] But, good movie if you can get through the [...] bits and pieces cut out,” remarked a third (Ugarnyj Zaškvár, 2021).

TV series are equally susceptible to unauthorized uploading. YouTube offers episodes for rental from several seasons of the ABC reality series, *The Bachelor*. Users occasionally upload episodes, as appeared to be the case with a compilation of three episodes from season 25 uploaded in January 2021. Titled “the bachelor season 25 episode 3 full episodes,” it was still available on YouTube seven months later, after 36,000 views. This compilation of episodes was masked through a combination of reediting and reframing. Similar to the *Homefront* upload, scenes were reedited to include briefly repeated shots. Some shots were also reframed within each scene into grainy, extreme close-ups. As such, faces were often cropped, showing only part of the hair and forehead, or the lips, or an eye. As one YouTube viewer wrote in the video comments section, “Why is the picture too close up to the point people’s faces are cut out of frame?” (Hainat, 2021).

Similar masking effects were applied to episodes of the courtroom reality series, *Judge Judy*, made available on YouTube from one account in February 2022 and had thousands of views. While the audio remained relatively untouched, the video was heavily reedited with repeating shots lasting only one or two seconds of plaintiffs and defendants, and intercut with shots of the series’ star—Judge Judith Sheindlin—that frequently lasted less than a second (Kanha Music Rasiya, 2022). *Judge Judy* episodes from the same period, but uploaded from another account, contained masking primarily through added extreme close-ups like those in the compiled episodes of *The Bachelor* (ExPol, 2022).

It is debatable whether such altered works can be detected by Content ID’s automated processes (Douglas, 2018; Engstrom and

Feamster, 2017; Herrman, 2013; Seidler, 2019). As Regina Zernay and Ryan Hagemann explain, “[w]hat constitutes a ‘match’ depends on how the software has been programmed. Factors such as the types and percentage of similarities serve as parameters for the software’s evaluation” (2017, p. 8). In their work on automated anti-piracy systems, Evan Engstrom and Nick Feamster note that “because all filtering technologies rely on examining and analyzing aspects of target files [...] all filtering tools can be evaded through basic manipulation of the file” (2017, p. 17-18). However, when videos of unauthorized content are posted with accurate titles, as with the episodes of *The Bachelor* considered here, it is hard to believe they would evade detection by even the most elementary systems. Although the visual and auditory alterations to an uploaded video may deceive anti-piracy AI, they may also make it so drastically different from the original that—even when detected—it would not appear to endanger the original’s commercial value. We could thus say that anti-piracy AI impacts the techniques and aesthetics of digital piracy regardless of whether the altered files pose any significant challenge to the system’s algorithms and machine-learning abilities.

CONCLUSION

Lev Manovich posits that although the current online media environment of frequent sampling, alteration, and recirculation of existing works may seem entrenched, it will be supplanted by other forms of media production and circulation. “If we [...] try now to develop a better historical and theoretical understanding of the remix era and the technological platforms that enable it, we will be in a better position to recognize and understand the new era that will inevitably replace it,” says Manovich (2015, p. 124). The creative piracy examined here presents an intriguing and overlooked category of production and circulation. Unlike other types of remix that seek to transform existing material to create and extract new meanings, aesthetic alterations for piracy typically seek to preserve the core of the existing text while altering superficial aspects. With copyright violators finding new ways to mask, and with machine-learning anti-piracy systems becoming more and more sophisticated in uncovering altered works, the resulting cat-and-mouse game may impact how cultural works are made, while changing understandings of what constitutes fair use.

A study of mashup producers, for example, found that automated content recognition, on platforms such as YouTube and Facebook, can significantly affect how users make works from copyrighted material. Arguing that “platforms’ content moderation currently supersedes, in practice, the legal exceptions intended to allow for particular uses” such as transformative creations and parody, the study’s authors noted that “platforms’ detection procedures also have an *indirect* impact on the mashup producers, who utilise a wide variety of workarounds to bypass detection, which in turn significantly affects their creative decisions and distribution strategies” (Brøvig-Hanssen and Jones, 2021). While that study focuses on users who intend to create new, transformative works from existing copyrighted material, it points to the sophistication, reach, and impact of automated anti-piracy systems. These systems flag and remove user-generated content that may be considered fair use under most understandings of contemporary copyright law, yet the commercial and social power of these platforms makes legitimate remix producers wary of challenging the system for fear of losing their accounts and followers. In such a climate, the relationship between these systems and material that is not meant to be transformative, but merely masked, can raise important questions about what constitutes transformative change of a cultural work, thus opening the door to new rulings on fair use. Although AI systems can be trained to detect most online copies of a work, they may not recognize some of them because of significant changes to the content. Such copies may meet the creativity threshold that justifies their legal protection as fair use. Whereas fair use rulings typically consider the intent and nature of the use of the copyrighted material, here they could also take into account the ability to deceive sophisticated machine-learning systems. In essence, the AI system perceives the altered copy as being fundamentally different from the original, and even unrecognizable. If, on the other hand, these anti-piracy systems are sophisticated enough to identify even significantly distorted, degraded, or otherwise altered copies of copyrighted works, one could infer from the continuing presence of some of this material that rights holders may feel it lies outside the bounds of narrative intelligibility and market relevance. Though such altered copies may continue to tell a story, it is no longer the protected story of the original, but rather a story about the limits to aesthetic alteration and meaning in online social media environments.

REFERENCES

- Berry, David M., *Copy, Rip, Burn: The Politics of Copyleft and Open Source*, London, Pluto Press, 2008.
- Brøvig-Hanssen, Ragnhild and Ellis Jones, “Remix’s retreat? Content moderation, copyright law and mashup music”, *New Media & Society*, 2021.
- Burgess, Jean and Joshua Green, *YouTube: Online Video and Participatory Culture*, New York, Polity, 2009.
- Candance Celeste Stmarie, “Dwayne Johnson movie 2021 – New action movies full English 2021 – Best action movie HD – movie 2021”, *YouTube*, 3 December 2021. <<https://www.youtube.com/watch?v=eWaaPcRzwdU>>.
- Douglas, Nick, “It’s supposed to look like shit: The internet ugly aesthetic”, *Journal of Visual Culture*, vol. 13, n° 3, 2014, p. 314-339.
- Douglas, Nick, “You can’t fool YouTube’s copyright bots”, *Lifehacker*, 24 January 2018, <<https://lifehacker.com/you-cant-fool-youtubes-copyright-bots-1822174263>>.
- Edwards, Dustin W., “Circulation gatekeepers: Unbundling the platform politics of YouTube’s Content ID”, *Computers and Composition*, vol. 47, 2018, p. 61-74.
- Eichhorn, Kate, *Adjusted Margin: Xerography, Art, and Activism in the Late Twentieth Century*, Cambridge, MIT Press, 2016.
- Engstrom, Evan and Nick Feamster, “The limits of filtering: A look at the functionality & shortcomings of content detection tools”, *Engine*, March 2017, <<https://www.engine.is/the-limits-of-filtering>>.
- ExPol, “Judge Judy episodes 7000 best amazing cases season 2022”, *YouTube*, 13 February 2022, <<https://www.youtube.com/watch?v=7wpDVmPUgFo>>.
- Frangoul, Anmar, “With over 1 billion users, here’s how YouTube is keeping pace with change”, *CNBC*, 14 March 2018, <<https://www.cnbc.com/2018/03/14/with-over-1-billion-users-heres-how-youtube-is-keeping-pace-with-change.html>>.
- Freund, Katharina, “‘Becoming a part of the storytelling’: Fan vidding practices and histories”, In: Paul Booth (ed.), *A Companion to Media Fandom and Fan Studies*, Oxford, Wiley, 2018.
- Google, “How Google Fights Piracy”, November 2018, <<https://www.google.com/url?sa=t&rc=t=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjJrJSHnZjyAhXHEVkJFHx-UDIMQFnoECA8QAw&url=https%3A%2F%2Fwww.blog.google%2Fdocuments%2F27%2F&usg=AOvVaw2yYhKtui1ctBduK5ZhXQzF>>.
- Gray, Joanne Elizabeth, *Google Rules: The History and Future of Copyright under the Influence of Google*, Oxford, Oxford University Press, 2020.
- Hainat, “The Bachelor season 25 episode 3 full episodes”, *YouTube*, 21 January 2021, <<https://www.youtube.com/watch?v=tvVEvmhLqro>>.
- Herrman, John, “There are still hundreds of pirated movies on YouTube, and they’re hiding in plain sight”, *BuzzFeedNews*, 18 July 2013, <<https://www.buzzfeednews.com/article/jwherrman/there-are-still-hundreds-of-pirated-movies-on-youtube-and-th>>.
- Jenkins, Henry, *Fans, Bloggers, and Gamers: Exploring Participatory Culture*, New York, New York University Press, 2006.
- Jenkins, Henry, Sam Ford and Joshua Green, *Spreadable Media: Creating Value and Meaning in a Networked Culture*, New York, NYU Press, 2013.

- Johns, Adrian, *Piracy: The Intellectual Property Wars from Gutenberg to Gates*, Chicago, University of Chicago Press, 2011.
- Kanha Music Rasiya, "JUDGE JUDY EPISODE 9377 amazing cases season 2022 full episode", *YouTube*, 17 February 2022, <<https://www.youtube.com/watch?v=F0TV46WbOsI>>.
- King, David, "Latest content ID tool for YouTube", *Google Official Blog*, 15 October 2007, <<https://googleblog.blogspot.com/2007/10/latest-content-id-tool-for-youtube.html>>.
- Lessig, Lawrence, *Remix: Making Art and Commerce Thrive in the Hybrid Economy*, New York, Penguin Books, 2008.
- Ligon, Sarah, "AI can create art, but can it own copyright in it, or infringe?", *Lexis-Nexis*, 28 February 2019, <<https://www.lexisnexis.com/lexis-practical-guidance/the-journal/b/pa/posts/ai-can-create-art-but-can-it-own-copyright-in-it-or-infringe>>.
- Manovich, Lev, "Remix strategies in social media", in Eduardo Navas, Owen Gallagher, and xtine burrough (eds.), *The Routledge Companion to Remix Studies*, New York, Routledge, 2015.
- Moore, Thurston, *Mix Tape: The Art of Cassette Culture*, New York, Universe, 2005.
- Navas, Eduardo, *Remix Theory: The Aesthetics of Sampling*, New York, Springer, 2012.
- Netflix, "Global Top 10" <<https://top10.netflix.com/>>.
- Red Points, "Why Red Points?", 2023, <<https://www.redpoints.com/why-red-points/>>.
- Seidler, Ellen, "YouTube's content ID easily fooled", *Voxindie*, 5 January 2019, <<https://www.voxindie.org/youtubes-content-id-easily-fooled/>>.
- Sobel, Benjamin, "Artificial intelligence's fair use crisis", *Columbia Journal of Law & the Arts*, vol. 45, 2017, p. 46-97.
- Trendacosta, Katherine, "Unfiltered: How YouTube's Content ID discourages fair use and dictates what we see online", *Electronic Frontier Foundation*, 2020, <<https://www.eff.org/unfiltered>>.
- Vézina, Brigitte and Brent Moran, "Artificial intelligence and creativity: Why we're against copyright protection for AI-generated output", *Creative Commons*, 10 August 2020, <<https://creativecommons.org/2020/08/10/no-copyright-protection-for-ai-generated-output/>>.
- YouTube Creators, "YouTube Content ID", *YouTube*, 28 September 2010, <<https://www.youtube.com/watch?v=9g2U12SsRns>>.
- Wagener, Albin, "Mèmes, gifs et communication cognitivo-affective sur Internet. L'émergence d'un nouveau langage humain", *Communication*, vol. 37, n° 1, 2020.
- Zaškar, Ugarnyj [Угарный Зашквар], "King of special forces – movie powerful action 2021 full length English latest new best action movies", *YouTube*, 9 July 2021, <<https://www.youtube.com/watch?v=uzHrR5slXdc>>.
- Zernay, Regina and Ryan Hagemann, "ACES in the hole? Automated copyright enforcement systems and the future of copyright law", *The Niskanen Center*, 6 June 2017.

ALGORITHMES DE RECOMMANDATION

DISPOSITIF DE LA GOUVERNEMENTALITÉ

ALGORITHMIQUE MENAÇANT LES DROITS HUMAINS ?

Lucia Flores Echaiz¹

L'intelligence artificielle (IA) affecte le champ culturel et médiatique autant dans la production que dans la distribution et la réception de contenus. Par rapport à la production, des systèmes d'IA peuvent générer des œuvres artistiques de façon quasi autonome ou être utilisés par des artistes dans leurs démarches de création (Christie's, 2018; McCarthy, 2016; Artaigallery, n.d.). Des systèmes d'IA sont également utilisés dans plusieurs étapes du processus journalistique, à savoir la collecte des données, l'extraction des informations pertinentes et l'écriture automatisée (Hu *et al.*, 2019; Aguirre Hernando, 2017; Cabra, 2017; Moses, 2017; Lemelshtrich Latar, 2015). Pour ce qui est de la distribution et de la réception du contenu culturel et médiatique, diverses plateformes utilisent des systèmes d'IA pour offrir des recommandations personnalisées à la personne utilisatrice. Ce chapitre se concentre sur ce dernier type d'algorithmes.

Les algorithmes de recommandation sont effectivement employés par plusieurs plateformes culturelles telles que Spotify et Netflix pour montrer aux personnes utilisatrices du contenu personnalisé. Les algorithmes de recommandation jouent un rôle important dans les réseaux sociaux. En se basant sur la géolocalisation, l'historique de navigation, les achats effectués, l'analyse sémantique des publications et autres facteurs,

1. Avocate en droit criminel et candidate à la maîtrise en droit à l'Université du Québec à Montréal (UQAM).

les algorithmes de personnalisation privilégient des publications adaptées aux besoins et aux goûts prédits des personnes utilisatrices (Hu *et al.*, 2019, p. 35-36). Sur les réseaux sociaux, ces publications sont très variées et englobent des nouvelles sur l'actualité, des annonces pour des événements et des publicités pour une panoplie de produits et services, notamment pour des albums de musique, des films, des livres et diverses marchandises. Notons également que plusieurs moteurs de recherche, tels que Google, emploient des algorithmes de personnalisation du contenu, ce qui peut amener des individus faisant des recherches identiques à recevoir des résultats différents (Bezemek, 2020). Ces plateformes dites intermédiaires « contrôlent les canaux de communication que les éditeurs de journaux avaient jadis et modulent ainsi le type de nouvelles (et de publicités) qui rejoint les personnes utilisatrices² » (Nechushtai et Lewis, 2019, p. 299).

Bien qu'une recommandation d'un morceau de musique ou de marchandise ne semble pas *a priori* revêtir une grande importance philosophique ou politique, cette tendance à la personnalisation algorithmique a des répercussions sur les individus d'une façon à laquelle il convient de s'attarder. En effet, ces algorithmes ont des conséquences sur la capacité des individus à connaître une gamme d'informations et d'expériences, ce qui soulève des questions d'intérêt public, plus particulièrement des interrogations sur leurs effets sur les droits humains.

À la suite d'une présentation du problème, à savoir les algorithmes de recommandation (1), cet article se penchera sur la notion de gouvernamentalité algorithmique (2). Cette notion sert à illustrer comment les algorithmes qui se nourrissent de données massives ont des effets – voire du pouvoir – sur plusieurs sphères de nos vies. Ainsi, nous pourrions examiner plus en profondeur dans quelle mesure les algorithmes de recommandation peuvent menacer les droits humains suivants : la liberté d'expression et la liberté d'opinion (2.1), la vie privée (2.2) et le droit à l'égalité (2.3). Deux remarques s'imposent. Premièrement, nous nous pencherons uniquement sur le régime international des droits humains. Deuxièmement, cet article ne vise pas à affirmer qu'une plateforme en particulier viole un ou plusieurs droits humains et qu'ainsi un groupe d'individus auraient certainement gain de cause devant les tribunaux. Nous visons plutôt à continuer les réflexions sur les effets que ces algorithmes ont sur les fondements des droits humains mentionnés.

2. Traduction libre de l'anglais vers le français.

1. ALGORITHMES DE RECOMMANDATION

Les algorithmes jouent un rôle incontournable dans la façon dont les individus accèdent à du contenu culturel (Morris *et al.*, 2021 ; O'Dair et Fry, 2020 ; Lury et Day, 2019 ; Roberge, Jamet et Rousseau, 2018 ; Wilson-Barnao, 2017 ; Striphas, 2015). Vu que « [l]es industries culturelles forment un domaine où l'incertitude sur la qualité des produits, le risque commercial et la surcharge d'informations [...] sont omniprésents » (Ménard, 2014, p. 70), il n'est pas surprenant que des compagnies et autres acteurs se soient mis à employer de tels outils pour personnaliser l'offre culturelle qui est montrée aux individus. Cela a amené certaines personnes chercheuses à utiliser le terme « culture algorithmique » (Varela et Kaun, 2019 ; Wilson-Barnao, 2017 ; Hallinan et Striphas, 2016 ; Striphas, 2015). Plus précisément, Hallinan et Striphas reconnaissant que le concept de culture est polysémique, énoncent que la notion de culture algorithmique renvoie aux systèmes informatiques qui sélectionnent, classifient et hiérarchisent des personnes, des endroits, des objets et des idées ainsi qu'aux conduites, habitudes et formes d'expressions qui découlent en raison de ces processus computationnels (Hallinan et Striphas, 2016, p. 119). Il faut retenir que ces algorithmes ont pour but de montrer à la personne utilisatrice ce qui devrait être le plus pertinent pour elle. Un algorithme de recommandation est donc un système qui « vise à aider des personnes utilisatrices à trouver des biens auxquels elles sont intéressées en leur fournissant des recommandations basées sur leurs comportements » (Nell, 2014, p. 40)³. L'expression (algorithmes de recommandation) ne réfère donc pas à une technique en particulier. Il existe, au contraire, une panoplie de techniques utilisées à des fins de découvrabilité de contenu. Cela étant dit, il est possible de classifier les algorithmes de recommandation en deux catégories principales, à savoir les techniques de filtrage collaboratif et les approches basées sur le contenu (Lury et Day, 2019 ; Nell, 2014 ; Park *et al.*, 2012 ; Burke *et al.*, 2011 ; McEnnis et Cunningham, 2007)⁴.

Le filtrage collaboratif (*collaborative filtering*) repose sur l'idée que, si une personne aime la même chose qu'une autre personne, il y a de fortes chances que ces personnes aient d'autres goûts en commun. Autrement dit, ces algorithmes s'appuient sur des données reflétant les préférences des

3. Traduction libre de l'anglais vers le français.

4. Toutes les personnes autrices citées ici classifient les algorithmes de recommandation en ces deux catégories, mis à part Nell, qui en mentionne d'autres catégories, notamment les approches démographiques dont nous feront brièvement mention.

personnes utilisatrices et prédisent ce que ces personnes vont aimer en se basant sur des similarités avec les données d'autres personnes utilisatrices (Lury et Day, 2019, p. 22 ; Karakayali *et al.*, 2018, p. 7 ; Ménard, 2014, p. 73-74). Ces données de rétroaction peuvent être explicites ou implicites (Rodriguez, 2020, p. 12 ; Nell, 2014, p. 39 ; Johnson, 2014, p. 1-2 ; Lekakos et Giaglis, 2006, p. 412). Lorsque, par exemple, un individu signale qu'il aime une chanson en cliquant sur un bouton « j'aime », cela représente une donnée de rétroaction explicite. Les données de rétroaction implicites englobent quant à elles le fait qu'un individu écoute une chanson, le temps d'écoute, la fréquence à laquelle il écoute la même chanson, etc. Notons également que les systèmes de filtrage collaboratif sont surtout performants lorsqu'on a accès à un grand nombre de données (Maillet, 2009, p. 20). Ainsi, un problème avec ces techniques surgit au début de l'utilisation, soit lorsqu'il n'y a pas suffisamment de données sur les nouvelles personnes utilisatrices. Il s'agit du problème du démarrage à froid (Nell, 2014, p. 41 ; Burke *et al.*, 2011, p. 16 ; Maillet, 2009, p. 20). La plateforme Last.fm représente un exemple de filtrage collaboratif ; il s'agit d'une plateforme de musique qui personnalise ce qui est transmis à la personne utilisatrice en se basant sur ce qu'elle a écouté et en comparant cela au comportement musical d'autres individus (Lury et Day, 2019, p. 22 ; Karakayali *et al.*, 2018 ; Ménard, 2014, p. 74)⁵. Dans la même lignée, nous pouvons penser à la compagnie Amazon et à leur « les clients qui ont aimé ce que vous venez d'acheter ont aussi aimé [...] » (Beuscart *et al.*, 2019, p. 22).

Ensuite, il y a les approches basées sur le contenu (Lury et Day, 2019 ; Castagnos *et al.*, 2014 ; Nell, 2014 ; Burke *et al.*, 2011). Ces systèmes fonctionnent à partir des propriétés d'un contenu que la personne utilisatrice a aimé pour recommander un autre contenu ayant des propriétés similaires. La plateforme Pandora Internet Radio utilise un algorithme de la sorte (Lury et Day, 2019, p. 22)⁶. Contrairement aux systèmes de filtrage collaboratif, les approches basées sur le contenu ne souffrent pas du problème du démarrage à froid, étant donné que, dès les premières données liées au comportement de la personne utilisatrice, il est possible d'en tirer des propriétés et ainsi de recommander de nouveaux éléments culturels ayant celles-ci. De plus, puisque les recommandations se basent sur les propriétés des éléments, ces systèmes sont davantage transparents et explicables (Maillet, 2009, p. 21-22).

5. Cela était le cas du moins jusqu'en 2019.

6. *Ibid.*

Également, bien que, dans les travaux les plus récents consultés sur le sujet, les approches démographiques ne soient pas mentionnées en tant que catégorie à part, il nous paraît néanmoins important d'en faire mention. Ces approches consistent à faire des corrélations entre les goûts des personnes utilisatrices et leurs informations démographiques, telles que le genre, l'âge, le niveau d'éducation, etc. (Nell, 2014, p. 42 ; Adomavicius et Tuzhillin, 2005, p. 20-19). Lorsqu'il n'y a pas suffisamment de données explicites dans le filtrage collaboratif, il est possible de considérer des personnes utilisatrices en tant que « similaires » en se basant sur des facteurs démographiques (Adomavicius et Tuzhillin, 2005, p. 19). Les approches basées sur le contenu peuvent, elles aussi, utiliser des éléments démographiques des personnes utilisatrices en liant ces derniers à des propriétés des éléments (Johnson, 2014, p. 1)⁷.

En effet, les trois techniques décrites ci-dessus ne doivent pas être comprises comme mutuellement exclusives. La plateforme Netflix, par exemple, utilise une formule hybride (Lury et Day, 2019, p. 22 ; O'Dair et Fry, 2020, p. 74 ; Bucher, 2018, p. 48)⁸. Ainsi, il est fort probable qu'une plateforme emploie plusieurs sortes d'algorithmes en même temps et que ceux-ci évoluent en temps réel (Beuscart *et al.*, 2019, p. 19 ; O'Dair et Fry, 2020, p. 74).

Prenons l'exemple de la plateforme musicale Spotify. Celle-ci emploie des algorithmes de recommandation afin de proposer de nouvelles chansons à la personne utilisatrice et construire différentes listes de lecture personnalisées (Spotify, 2019). Cette personnalisation peut faire en sorte qu'un artiste soit visible pour une personne et pratiquement exclu pour une autre (O'Dair et Fry, 2020, p. 73). Christopher Johnson, alors gestionnaire de l'équipe en apprentissage automatique à Spotify, expliquait que, parmi les motivations pour utiliser ces algorithmes, il y a le fait d'accroître la satisfaction des personnes utilisatrices et de générer davantage de profits pour l'entreprise (Johnson, 2014, p. 1). Notons également que les données explicites peuvent être parfois difficiles à obtenir et qu'ainsi les méthodes tirant des corrélations sur un ensemble de données de rétroaction implicites

7. Nous supposons donc que la majorité de la littérature ne considère pas les approches démographiques comme une catégorie à part puisque l'utilisation des éléments démographiques peut être faite autant avec les approches basées sur le contenu qu'avec les approches de filtrage collaboratif.

8. Lury et Day énoncent qu'il s'agit d'une combinaison de techniques de filtrage collaboratif et de techniques basées sur le contenu. O'Dair et Fry, s'appuyant sur Bucher, énoncent quant à eux que la plateforme utilise probablement 5 ou 10 algorithmes différents.

peuvent être parfois plus intéressantes (Johnson, 2014, p. 3). Spotify se concentre donc davantage sur les données implicites et utilise principalement la technique du filtrage collaboratif (Huq, 2019 ; Johnson, 2014). Plus précisément, Spotify emploie un modèle probabiliste (*logistic matrix factorization*) ayant pour but de prédire la probabilité qu'une personne utilisatrice préfère un morceau de musique (Huq, 2019 ; Johnson, 2014). Cela étant dit, il n'est pas possible de savoir avec exactitude le fonctionnement des algorithmes de Spotify. Cette opacité, à la fois technique – surnommée le problème de la boîte noire (O'Dair et Fry, 2020 ; Hallinan et Striphos, 2016 ; Pasquale, 2015) –, mais également corporative, est un obstacle récurrent entourant les systèmes d'IA.

D'après les informations disponibles, il semble que Spotify se base uniquement sur les données recueillies sur leur site. Cela semble également être le cas de Netflix qui affirme ne pas se servir des données d'autres applications (Varela Martinez, 2017, p. 16). Par contre, la logique d'autres algorithmes de recommandation est souvent moins transparente et davantage omniprésente :

The amount of time an individual spends looking at a particular object, how often they respond to a page or friend, what they like, share or comment on, and their history of interaction with similar posts (including what they have hidden or reported) [can all enable] the algorithm to work more effectively (Wilson-Barnao, 2017, p. 561).

De plus, les discussions entourant les algorithmes de recommandation font souvent mention de la notion de bulles de filtre (*filter bubbles*) (Liu *et al.*, 2021 ; Bezemek, 2020 ; Karakayali *et al.*, 2018 ; Flaxman *et al.*, 2016 ; Gottron et Schwagereit, 2016 ; Pariser, 2011). La notion désigne le fait qu'avec les algorithmes de personnalisation chaque personne serait dans une bulle d'information, qui ne serait pas la même que celle d'un autre individu. Ainsi, les limites des données d'une personne représenteraient les limites de son monde (Bezemek, 2020). Cette notion vise à illustrer le pouvoir d'exclusion et d'inclusion que les algorithmes de recommandation ont sur nos vies (Karakayali *et al.*, 2018, p. 5). Afin de continuer d'examiner de façon critique les effets de ces algorithmes, nous croyons pertinent d'inclure dans notre réflexion la notion de gouvernementalité algorithmique.

2. GOUVERNEMENTALITÉ ALGORITHMIQUE

La gouvernementalité algorithmique désigne « un certain type de rationalité (a)normative ou (a)politique reposant sur la récolte, l'agrégation

et l'analyse automatisée de données en quantité massive de manière à modéliser, anticiper et affecter par avance les comportements possibles» (Rouvroy et Berns, 2013, p. 173). La notion peut être également comprise comme l'idéologie entourant le phénomène des données massives (*big data*) (Rouvroy, 2014). Les effets concrets de cette gouvernamentalité algorithmique peuvent aller d'un refus de crédit bancaire à une publicité personnalisée sur les réseaux sociaux.

Un des éléments les plus importants de la théorie est le rapport au réel que la gouvernamentalité algorithmique amène (Rouvroy et Stiegler, 2015, p. 114; Rouvroy, 2016a, p. 11-16). Dans la rationalité algorithmique, nul besoin d'interprétations individuelles ou collectives sur le monde. Le réel – par les données devenant des faits objectifs – parle de lui-même et un accès immédiat à lui s'opère (Rouvroy, 2014, p. 8; Rouvroy, 2016a, p. 12). Ainsi, les profils algorithmiques jouissent d'une aura d'impartialité et d'objectivité (Rouvroy et Berns, 2010, p. 92; Rouvroy, 2016a, p. 11). Toujours selon cette rationalité algorithmique, la réalité devient enfin saisissable et saisie : « l'idée étant que le "savoir" ou la "vérité" ne seraient plus construits, mais toujours déjà là, immanents aux banques de données » (Rouvroy, 2014, p. 9). Il est pourtant clair que ce rapport au réel est, au contraire de sa prétention, très idéologique. Ce rapport sous-tend une naturalisation des données. En effet, le processus informatique décontextualise les données, faisant qu'elles » n'appara[issent plus] comme le résultat des rapports de pouvoir » (Rouvroy, 2014, p. 8).

La gouvernamentalité algorithmique s'intéresse à nous, non pas en tant que sujets, mais en tant qu'agréats des données (Rouvroy et Berns, 2010, p. 97). Elle vise à cibler les besoins, les envies et les préférences de chaque individu, mais il ne s'agit pas d'une forme d'individuation valorisant la singularité de chaque personne. Il s'agit plutôt d'une hypersegmentation des offres commerciales (Rouvroy et Berns, 2013, p. 176). Il n'y a pas de place pour les rêves, pour ce qu'on pourrait devenir ou pour la façon dont on perçoit singulièrement notre histoire (Rouvroy, 2014, p. 3).

Tout cela a des effets sur nous et sur nos potentialités. La gouvernamentalité algorithmique fait en sorte que nous sommes là où on est attendus, en conditionnant ce que nous achèterons, nous lirons, nous regarderons et où nous irons (Rouvroy, 2014, p. 6). Nous sommes en quelque sorte « enfermés dans une bulle immunitaire hyper-personnalisée » (Rouvroy, 2014, p. 6). Rouvroy énonce que ce profilage colle aux individus comme une seconde peau (Rouvroy et Stiegler, 2015, p. 123).

La gouvernementalité algorithmique peut avoir un air d'émancipation, étant distincte des formes de biopouvoir et de gouvernementalité disciplinaire (Rouvroy, 2016b, p. 34). Au lieu de nous imposer une norme universelle qui punirait les transgressions à celle-ci, la norme dans la gouvernementalité algorithmique est flexible. La norme qui nous est imposée est le profil algorithmique qui délimite nos possibilités, profil construit par les traces numériques de notre personne et évoluant au fil du temps (Rouvroy et Berns, 2013, p. 173). Autrement dit, il ne s'agit « plus d'exclure ce qui sort de la moyenne, mais d'éviter l'imprévisible, de faire en sorte que chacun soit véritablement lui-même » (Rouvroy et Berns, 2013, p. 172-173).

Le piège dans cette apparente normativité émancipatrice est que « le gouvernement algorithmique “crée” une réalité au moins autant qu'il l'enregistre » (Rouvroy et Berns, 2013, p. 183). En effet, nous ne sommes pas gouvernés par un reflet authentique de nous-mêmes. Il y a des choses qui échappent à la numérisation et qui sont très importantes culturellement et politiquement, telles que les choses qui n'ont pas été actualisées ou effectuées dans le passé, soit des « projets inaboutis [ou des] utopies non réalisées » (Rouvroy et Stiegler, 2015, p. 119). Également, malgré l'aspiration de la gouvernementalité algorithmique, l'avenir n'est pas numérisable non plus.

3. RÉPERCUSSIONS SUR LES DROITS HUMAINS

Nous expliquerons dans cette partie les tensions entre la gouvernementalité algorithmique, présente dans les algorithmes de recommandation, et les fondements de certains droits humains. Nous avons choisi de nous concentrer sur la liberté d'opinion et la liberté d'expression, le droit à la vie privée et le droit à l'égalité, car il nous semble que ce sont les principaux droits menacés par les algorithmes de recommandation. Notons aussi que, bien que le régime international des droits humains s'applique principalement aux acteurs étatiques, selon les principes Ruggie, les compagnies privées doivent également les respecter⁹.

9. *Principes directeurs relatifs aux entreprises et aux droits de l'homme : mise en œuvre du cadre de référence « protéger, respecter et réparer » des Nations Unies*, Doc off Conseil des droits de l'homme, 17^e sess, annexe, Doc NU A/HRC/17/31 (2011) 7.

3.1 Liberté d'opinion et liberté d'expression

La liberté d'opinion et la liberté d'expression sont étroitement liées. La *Déclaration universelle des droits de l'homme* (DUDH) prévoit les deux à l'article 19 : « Tout individu a droit à la liberté d'opinion et d'expression, ce qui implique le droit de ne pas être inquiété pour ses opinions et celui de chercher, de recevoir et de répandre, sans considération de frontières, les informations et les idées par quelque moyen d'expression que ce soit »¹⁰. Le *Pacte international relatif aux droits civils et politiques* (PIDCP) prévoit, également à l'article 19, que « [n]ul ne peut être inquiété pour ses opinions »¹¹ et que toute personne a droit à la liberté d'expression, ce qui comprend « la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce [...] »¹². Le droit à la liberté d'opinion implique le droit de former une opinion sans coercition¹³. Ces deux libertés sont à la fois nécessaires au développement de toute personne et essentielles pour une société libre et démocratique¹⁴. De plus, bien que l'article 19 paragraphe 3 du PIDCP dispose que des restrictions peuvent s'appliquer à la liberté d'expression, il ne peut y avoir d'exceptions ou de limitations à la liberté d'opinion.

Deux dimensions se dégagent de ces dispositions : d'un côté, il y a la liberté de chercher et recevoir des informations et de former une opinion et, de l'autre côté, il y a la liberté de répandre des informations et des opinions (de Zayas et Roldán Martín, 2012). Bien que les systèmes d'IA aient des répercussions sur ces deux dimensions (Hu *et al.*, 2019), nous nous concentrons uniquement sur la première dimension, étant donné que les algorithmes de recommandation ont des répercussions principalement sur celle-ci.

La fonction avouée des algorithmes de recommandation implique l'inclusion et l'exclusion de certains contenus. Ainsi, il n'est pas surprenant que plusieurs s'inquiètent des effets négatifs de ces algorithmes sur les libertés d'opinion et d'expression (Eskens *et al.*, 2017 ; Hu *et al.*, 2019 ;

10. *Déclaration universelle des droits de l'homme*, Rés AGA 217A (III), Doc off AG NU, 3^e sess, supp n° 13, Doc NU A/810 (1948) 71, art. 19.

11. *Pacte international relatif aux droits civils et politiques*, 999 RTNU 171, 19 décembre 1966 (entrée en vigueur : 23 mars 1976) [PIDCP], art. 19, au paragr. 1.

12. *Ibid.*, au paragr. 2.

13. Rapport du *Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression*, Doc off AG NU, 73^e sess, Doc NU A/73/348 (2018), p. 12.

14. *Observation générale n° 34 : Article 19 : Liberté d'opinion et liberté d'expression*, Doc off Comité des droits de l'homme, 102^e sess, Doc NU CCPR/C/GC/34 (2011).

Access Now, 2018 ; Comité d'experts sur les intermédiaires d'Internet, 2017)¹⁵. L'enjeu principal des algorithmes de recommandation semble résider dans l'effet « bulle de filtres » mentionné précédemment. Il est en effet indéniable que les algorithmes influencent la sphère informationnelle des individus, affectant par le fait même leur liberté de recevoir des informations et leur capacité à former des opinions et des goûts. Ainsi, une personne se laissant guider par ces algorithmes n'aura pas facilement accès à des informations, à du contenu culturel et à des opinions n'ayant pas été prédits pour elle. Comme l'avance la théorie de la gouvernementalité algorithmique, ces systèmes informatiques délimitent en quelque sorte les dimensions de l'espace de possibilités qui se présentent à une personne.

Mentionnons toutefois deux interrogations nuancant (relativement) cet effet. D'une part, il est possible de se demander en quoi ces algorithmes diffèrent de l'édition de contenu préalable à l'ère algorithmique :

L'édition des contenus influence depuis longtemps la capacité d'une personne à se faire ses propres opinions : par exemple, les organes médiatiques mettent certaines informations à la une dans l'intention de façonner et d'influencer ce que chacun sait des événements de la journée¹⁶.

Ainsi, la question de l'effet algorithmique sur les libertés d'opinion et d'expression devient un effet de degré. Est-ce que les algorithmes de recommandation portent davantage préjudice à ces libertés que les techniques traditionnelles visant la découvrabilité de contenu ? Nous postulons qu'il y a des raisons de croire que oui. Le Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression explique certaines de celles-ci, à savoir le fait que les plateformes employant ces algorithmes exercent souvent un monopole rendant difficile pour les individus de s'y retirer, le fait que les recommandations apparaissent comme étant les plus pertinentes et objectives pour l'individu ainsi que la non-transparence de ces systèmes¹⁷.

Énoncer que les techniques algorithmiques modulent l'accès à du contenu culturel n'équivaut donc pas à nier que les techniques antérieures ne le faisaient pas non plus d'une autre façon, tout comme considérer que les algorithmes participent à une forme de gouvernementalité n'implique

15. Voir également le Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression, *supra* note 13.

16. *Ibid.*, p. 12.

17. *Ibid.*, p. 12-13.

pas que d'autres formes de gouvernementalité n'existaient pas au préalable. La norme culturelle que les individus sont poussés à suivre maintenant n'est plus universelle, mais bien individualisée. Le piège réside alors précisément dans le fait qu'il existe plusieurs éléments du profil qui nous est attribué auxquels il est possible de s'y identifier. Cela, jumelé à l'aura d'objectivité des systèmes algorithmiques, participe à sentir le profilage comme une seconde peau, comme le mentionne Rouvroy et qu'ainsi « ne pas vouloir être profilé revien[t] à ne pas se vouloir soi-même » (Rouvroy et Stiegler, 2015, p. 123).

D'autre part, afin de ne pas tomber dans une position technodéterministe, il est important de se demander quels sont les usages concrets de ces algorithmes ainsi que les perceptions des personnes utilisatrices sur ceux-ci. En effet, les effets des algorithmes « sont dépendants des contextes, évoluent avec les usages qui en sont faits et en fonction des données de leur environnement » (Beuscart *et al.*, 2019, p. 23). Les algorithmes ne sont pas les seuls facteurs qui influencent la façon dont les individus accèdent à du contenu culturel et médiatique. Par ailleurs, le phénomène des bulles de filtres semble difficile à quantifier empiriquement¹⁸. Cela étant dit, tel qu'énoncé par Eskens, certains groupes peuvent être davantage vulnérables aux effets de bulles de filtres ou peuvent l'être sur certains sujets (2020, p. 1121). Aussi, il est intéressant de noter qu'alors qu'une étude indiquait que l'algorithme de recommandation de Netflix influençait 75 % de l'activité de visionnement des personnes utilisatrices (Hallinan et Striphas, 2016, p. 130), une autre étude concernant Spotify indiquait que les écoutes autonomes représentaient les trois quarts des écoutes (Beuscart *et al.*, 2019, p. 42). Étant donné la non-généralisation des échantillons, les études empiriques, bien qu'elles soient pertinentes et nécessaires, demeurent pour le moment limitées quant au portrait général des utilisations et des effets de ces algorithmes. Enfin, bien que les algorithmes de recommandation ne déterminent pas complètement et automatiquement les opinions et les goûts des individus, il importe de reconnaître qu'ils jouent définitivement un rôle dans l'accès à de l'information et à du contenu culturel et qu'ils influencent ainsi la formation des opinions et des goûts, menaçant par le fait même les libertés d'expression et d'opinion.

18. Certaines personnes chercheuses considèrent que les preuves des effets de bulles de filtres sont contradictoires ou minimales (voir Eskens, 2020, p. 1121; Nechushtai et Lewis, 2019, p. 300; Eskens *et al.*, 2017, p. 280; Comité d'experts sur les intermédiaires d'Internet, 2017, p. 20).

3.2 Droit à la vie privée

Le droit à la vie privée est prévu à l'article 12 de la DUDH et à l'article 17 du PIDCP. Ceux-ci énoncent que « [n]ul ne sera l'objet d'immixtions arbitraires [...] dans sa vie privée, sa famille, son domicile ou sa correspondance » et que « [t]oute personne a droit à la protection de la loi contre de telles immixtions¹⁹. En 1988, dans son Observation générale n° 16, le Comité des droits de l'homme fournit des éclaircissements par rapport à ce droit et indique qu'il protège des immixtions provenant de l'État ou des personnes physiques ou morales²⁰. Bien qu'il ne définisse pas la notion de vie privée, le Comité énonce qu'« [é]tant donné que toutes les personnes vivent en société, la protection de la vie privée est nécessairement relative »²¹. Le droit à la vie privée n'est donc pas absolu (Rengel, 2014, p. 41)²². Par contre, pour qu'une atteinte soit légitime, elle doit, en plus d'être prévue par la loi, ne pas être arbitraire. Une atteinte non arbitraire est celle qui est conforme aux buts du PIDCP et qui est raisonnable dans les circonstances²³. Plus précisément, les informations que les autorités publiques peuvent réclamer doivent se limiter à celles « dont la connaissance est indispensable à la société »²⁴. Le degré du caractère nécessaire semble donc être assez élevé. Notons deux autres éléments pertinents dans cette Observation générale. Premièrement, le Comité interdit la surveillance électronique et l'interception, l'écoute et l'enregistrement des communications ; deuxièmement, le Comité fait mention de la notion des données personnelles en indiquant que tout individu devrait pouvoir déterminer si ses données sont stockées et à quelles fins²⁵.

19. *Déclaration universelle des droits de l'homme* ; *supra* note 10 ; *Pacte international relatif aux droits civils et politiques*, *supra* note 11. La différence réside dans l'ajout, dans le PIDCP, des mots « ou illégales » après le mot « arbitraires ».

20. *Observation générale n° 16 : Article 17 (Droit au respect de la vie privée)*, Doc off Comité des droits de l'homme 32^e sess, Doc NU HRI\GEN\1\Rev.1 (1988) [1994].

21. *Ibid.*, au para 7.

22. *Rapport du Rapporteur spécial sur le droit à la vie privée*, Doc off AG NU, Conseil des droits de l'homme, 40^e sess, Doc NU A/HRC/40/63 (2019), p. 4.

23. *Observation générale n° 16 : Article 17 (Droit au respect de la vie privée)*, *supra* note 20 ; *Rapport du Rapporteur spécial sur le droit à la vie privée*, *ibid.*, p. 5.

24. *Observation générale n° 16 : Article 17 (Droit au respect de la vie privée)*, *supra* note 20.

25. *Ibid.*

Plus récemment, le Haut-Commissaire des Nations Unies aux droits de l'homme indiquait que le respect de la vie privée d'une personne impliquait le droit d'un individu d'avoir

un espace dans lequel il peut s'épanouir, interagir et jouir d'une liberté en toute autonomie, une « sphère privée » dans laquelle il est libre d'interagir ou non avec d'autres personnes et peut échapper à l'intervention de l'État et à toute intervention excessive non sollicitée d'une tierce partie²⁶.

Ce lien étroit que le droit à la vie privée entretient avec l'autonomie personnelle est également souligné par le Rapporteur spécial sur le droit à la vie privée²⁷. Cette compréhension de la vie privée est distincte de la notion de protection des données personnelles mentionnée par le Comité des droits de l'homme. En effet, tel que nous l'avons énoncé dans une publication antérieure, la protection des données personnelles est une partie essentielle du droit à la vie privée, mais ne peut être considérée comme constituant son ensemble (Hu *et al.*, 2019, p. 45). Autrement dit, il est vrai que le contrôle qu'un individu possède quant au partage des informations le concernant occupe une place importante dans les discussions sur la vie privée et le numérique. Par contre, nous considérons qu'il est erroné de concevoir la vie privée qu'à travers ce prisme-là. La vie privée doit être également vue comme une place pour respirer (*breathing room*) essentielle à la détermination de soi (Cohen, 2013, p. 1920). Nous trouvons justes les propos suivants à ce sujet : « Le contrôle sur ses informations personnelles est un contrôle sur un aspect de l'identité qu'on veut projeter dans le monde et le droit à la vie privée est le fait d'être libre de contraintes déraisonnables dans la construction de sa propre identité » (Agre et Rotenberg, 1998, p. 7)²⁸. Mentionnons maintenant les principales menaces que les algorithmes de recommandation posent sur ces fondements.

26. *Rapport du Haut-Commissaire des Nations Unies aux droits de l'homme (Le droit à la vie privée à l'ère du numérique)*, Doc off AG NU, Conseil des droits de l'homme, 39^e sess, Doc NU A/HRC/39/29 (2018), p. 3.

27. *Rapport du Rapporteur spécial sur le droit à la vie privée* (2019), *supra* note 22, p. 4.

28. Traduction libre de l'anglais vers le français.

Tout d'abord, les algorithmes de recommandation débutent²⁹ tous avec la collecte des données sur les personnes utilisatrices. Dès ce moment-là³⁰, c'est-à-dire avant même de recevoir du contenu personnalisé, le droit à la vie privée est menacé dans sa dimension liée au contrôle des données personnelles.

D'une part, il existe une grande asymétrie informationnelle entre les plateformes et les personnes utilisatrices quant à cette récolte des données (Hu *et al.*, 2019, p. 48). Sans nier les efforts déployés pour réduire cette asymétrie, il ne nous semble pas exagéré d'affirmer que la plupart des personnes utilisatrices des plateformes comme Facebook, Twitter, Netflix, Spotify et YouTube, n'ont pas une connaissance significative quant aux traces numériques qu'elles laissent lors de leurs navigations et quant au traitement que ces plateformes en font³¹.

D'autre part, les contours de la notion de données personnelles deviennent flous avec les systèmes d'IA. Dubois et Gaullier énoncent que, pendant longtemps, les plateformes recueillant des informations sur les personnes utilisatrices, « notamment sur les pages visitées, les dates et heures de visite, les redirections effectuées de site en site, etc., [considérait que ces données étaient] anonymes, car non reliées à des informations permettant d'identifier nominativement les personnes concernées » (Dubois et Gaullier, 2017, p. 87). Néanmoins, le croisement des données considérées peu sensibles individuellement peut mener à un agrégat des données pouvant alors être considéré comme relevant du domaine personnel. Il est important de comprendre qu'il ne s'agit pas d'un effet pernicieux. Au contraire, la logique derrière ces systèmes est de recueillir le plus d'informations possible, notamment des données paraissant insignifiantes (Hu *et al.*, 2019, p. 46). Comment

29. Les systèmes d'IA alimentés par des données (massives) évoluent en temps réel, ce qui rend peut-être impropre l'usage du mot « début ». En effet, la collecte des données n'est pas un moment terminant avant le traitement des données. Il s'agit plutôt de processus qui ont lieu en même temps et qui interagissent sans cesse. Tout de même, à des fins de compréhension, il est possible de découper chronologiquement, bien qu'artificiellement, les différentes étapes du processus algorithmique.

30. Voir la description des trois « temps » de la gouvernamentalité algorithmique dans Rouvroy et Berns, 2013, p. 168-173.

31. Les raisons derrière ce manque d'information nous semblent être bien illustrées par l'explication tripartite de la notion d'opacité par Cofone (2019, p. 1438-1439). Premièrement, il y a l'enjeu des secrets commerciaux qui empêchent la transmission d'informations sur les systèmes déployés. Deuxièmement, le manque de littératie numérique chez le grand public ainsi que le manque de vulgarisation constituent des obstacles à la compréhension de ces processus. Troisièmement, lorsqu'il s'agit des systèmes d'IA complexes, l'explicabilité constitue un enjeu difficilement résoluble, rendant de ce fait davantage opaques ces systèmes.

un individu peut-il alors exercer du contrôle sur ses données si la finalité de leur collecte n'est pas connue ? Selon le Rapporteur spécial David Kaye, l'IA remet ainsi en question les principes sur lesquels se basent les normes internationales de protection des données, à savoir « les notions traditionnelles de consentement, de limitation des finalités et des utilisations, de transparence et de responsabilité » (Rouvroy, 2018, p. 427)³².

Il importe donc de retenir que, malgré les processus d'anonymisation, le risque d'être réidentifié persiste (Rouvroy, 2018, p. 425). Mentionnons l'exemple du concours « Netflix Prize », lancé en 2006, visant à récompenser l'équipe pouvant développer un système de recommandation qui soit plus performant que le système de la plateforme (Hallinan et Striphas, 2016, p. 118). Pour ce faire, Netflix a fourni un ensemble de données anonymisées constitué des notations de films provenant d'approximativement 500 000 personnes utilisatrices (Narayanan and Shmatikov, 2008, p. 10). Les chercheurs Narayanan et Vitaly ont alors démontré qu'il était possible de dé-anonymiser ces données, en les croisant avec des données non anonymes du site Internet Movie Database (IMDb) (Narayanan and Shmatikov, 2008, p. 11). Ainsi, pour eux, « la division naïve entre informations personnelles et informations non personnelles est une fausse dichotomie » (Narayanan and Shmatikov, 2008, p. 11)³³. Bien qu'*a priori* on pourrait avoir tendance à minimiser l'effet du statut public des notations de films de la part d'individus, leur étude souligne que des informations importantes sur les personnes utilisatrices, telles que leurs orientations politiques ou sexuelles ou d'autres informations n'ayant pas été rendues publiques par la personne dans la plateforme IMDb, peuvent être ainsi inférées (Narayanan et Shmatikov, 2008, p. 11, 16). À la suite du concours et de la publication de l'étude, une utilisatrice de Netflix a intenté un recours collectif contre la compagnie, alléguant une violation de sa vie privée (Singel, 2010)³⁴. Étant une femme lesbienne dont l'orientation sexuelle n'était pas dévoilée publiquement, la plaignante considérait, à notre sens raisonnablement, que des informations permettant d'inférer

32. *Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression*, *supra* note 13, p. 15.

33. Traduction libre de l'anglais vers le français.

34. La demande introductive d'instance peut être consultée au lien suivant : *Wired*, « Doe vs. Netflix », <https://www.wired.com/images_blogs/threatlevel/2009/12/doe-v-netflix.pdf>. Aucune décision ne fut rendue dans ce dossier, étant donné que l'affaire s'est close avec un règlement hors cour.

son orientation sexuelle (par exemple, le fait d'écouter beaucoup des films LGBTQ+) devraient être considérées comme personnelles et n'auraient pas dû être rendues publiques par Netflix (Singel, 2010). Remarquons cependant qu'il semble que les risques liés à la vie privée entourant le concours Netflix concernent la technique du filtrage collaboratif et non les approches basées sur le contenu (Narayanan et Shmatikov, 2008, p. 3 ; Calandrino *et al.*, 2011).

Ensuite, *après*³⁵ la collecte et le traitement des données arrive le profilage sur lequel les algorithmes de recommandation, peu importe l'approche choisie, se basent. Au-delà du risque de réidentification, les menaces à la vie privée persistent.

En prenant l'angle de la protection des données personnelles, nous sommes d'accord que « [l]a possibilité d'identifier nominativement un internaute n'est pas le seul indice de la qualification de donnée à caractère personnel » (Dubois et Gaullier, 2017, p. 87). Si des données liées aux caractéristiques ou aux comportements d'un individu *sont utilisées pour l'influencer – ce que les algorithmes de recommandation font* –, la protection des données à caractère personnel devrait alors s'appliquer (Groupe de travail « Article 29 » sur la protection des données, 2010, p. 11 ; Rodriguez, 2020, p. 16 ; Dubois et Gaullier, 2017, p. 87-88). Ainsi, la publicité ciblée et plus largement les recommandations basées sur des profils individuels sont des instances où les individus devraient avoir du contrôle sur leurs données.

Un point moins abordé dans les débats entourant la vie privée et le profilage consiste dans le fait que ce dernier vient miner l'autonomie personnelle ; autrement dit, le profilage algorithmique menace cet espace pour respirer (*breathing room*) (Cohen, 2013) devant être libre d'interférences extérieures. En effet, tandis que le droit à la vie privée vise à permettre le développement de l'identité personnelle en protégeant les individus de publicités ou de communications non voulues (Eskens, 2020, p. 1119), les algorithmes de recommandation renferment les individus dans des cases influençant leurs opinions et leurs goûts. Ce qui est particulièrement dérangent avec cette influence, c'est le fait que les individus ne sont pas conscients de ces processus et possèdent peu de moyens pour les défier. On peut ainsi comprendre davantage ce que Rouvroy cherchait à expliquer par la notion de gouvernementalité algorithmique et son effet sur le développement personnel. Alors que le droit à la vie

35. Voir notes 28, 29.

privée résonne avec l'idée de Rouvroy selon laquelle au 21^e siècle nous aimons nous comprendre comme des individus « en constante évolution, non finis et uniquement partiellement définis, ouverts à des nouveaux possibles³⁶ » (Rouvroy, 2016a, p. 36), les algorithmes de recommandation participent à une *fixation évolutive* des individus. Nous considérons que cet oxymore reflète bien la logique du profilage algorithmique : la norme culturelle qui nous est imposée évolue en temps réel à travers la collecte des données jamais terminée, mais elle renferme et fixe les limites du champ des possibles liées aux opinions et aux goûts d'un individu à un certain moment. Le droit à la vie privée à l'ère de la gouvernementalité algorithmique toucherait ainsi, selon nous, moins à un contrôle superficiel des données façonnant notre profil, mais à la capacité de ne pas suivre (et de ne pas être suivi par) les prédictions algorithmiques.

3.3 Le droit à l'égalité

En droit international, et plus largement dans les sociétés libérales (McMillan et Snelling, 2017, p. 71 ; Fredman, 2011, p. 1), l'égalité est une valeur centrale. Il est possible de la concevoir comme étant imbriquée dans tous les droits humains et dans tous les traités internationaux relatifs aux droits humains. L'article premier de la DUDH indique que « [t]ous les êtres humains naissent libres et égaux en dignité et en droits » et l'article 2 indique que toute personne jouit des droits proclamés « sans distinction aucune, notamment de race, de couleur, de sexe, de langue, de religion, d'opinion politique ou de toute autre opinion, d'origine nationale ou sociale, de fortune, de naissance ou de toute autre situation³⁷ ». Le PIDCP prévoit, de façon similaire, à l'article 2 que les États doivent garantir les droits mentionnés à tous les individus sans distinction et à l'article 26 que « [t]outes les personnes sont égales devant la loi et ont droit sans discrimination à une égale protection de la loi³⁸ ».

Soulignons aussi que l'égalité et la discrimination sont intimement liées, le droit à l'égalité impliquant le droit à la non-discrimination (Hu *et al.*, 2019 ; Bayefsky, 2015). De plus, le Comité des droits de l'homme indique que, tandis que l'article 2 du PIDCP est limité aux droits prévus dans le Pacte, l'article 26 interdit toute discrimination par les pouvoirs

36. Traduction libre de l'anglais vers le français.

37. Déclaration universelle des droits de l'homme, *supra* note 10.

38. Pacte international relatif aux droits civils et politiques, *supra* note 11.

publics³⁹. Bien que n'émanant pas des pouvoirs publics, les algorithmes de recommandation soulèvent des enjeux de discrimination.

La discrimination algorithmique est un sujet ayant fait couler beaucoup d'encre et qui a lieu dans différents domaines d'application⁴⁰. Les algorithmes de recommandation auxquels on s'intéresse n'en sont pas exempts. Ces algorithmes impliquent nécessairement des distinctions entre les personnes utilisatrices. Ainsi, il est possible de dire que la « discrimination, est à un certain niveau, inhérente au profilage : le but du profilage étant de traiter des personnes de façon différente » (Datta *et al.*, 2015, p. 105)⁴¹. Mentionnons un exemple significatif de discrimination à travers les algorithmes de personnalisation concernant des publicités liées à des services d'accompagnement professionnel pour des emplois. Les publicités ciblées étaient différentes pour les femmes et les hommes, les hommes recevant des publicités concernant des emplois mieux payés que ceux que les femmes recevaient (Datta *et al.*, 2015, p. 105).

Nous avons mentionné préalablement les approches démographiques et il nous semble que les stratégies de marketing plus traditionnelles se soient souvent basées sur de telles catégories pour faire des recommandations (Rodriguez, 2020, p. 8; Hallinan et Striphas, 2016, p. 123). Segmenter le public en des catégories pour vendre un produit peut sembler tout à fait normal et inoffensif. Par contre, au sujet de l'égalité de jouissance du droit à la liberté d'expression et la liberté d'opinion, il est possible de se demander si le fait de ne pas recevoir les mêmes informations que d'autres individus, en raison de notre appartenance à une caractéristique protégée, ne correspond pas à de la discrimination illicite. Il pourrait être alors intuitif de considérer que seuls les algorithmes de recommandation explicitement basés sur des facteurs démographiques, tels que le genre et l'âge, courent le risque d'être discriminatoires. En plus, il pourrait être rassurant de savoir que les facteurs démographiques sont de plus en plus délaissés dans les récentes approches d'algorithmes de recommandation (Hallinan et Striphas, 2016, p. 123). Par contre, le fait que les algorithmes ne se basent pas explicitement sur des motifs protégés n'exclut pas le fait que ceux-ci, avec les données, apprennent indirectement à faire des distinctions

39. *Observation générale n° 18: Non-discrimination*, Doc off Comité des droits de l'homme, 37^e sess, Doc NU HRI\GEN\1\Rev.1 (1989) [1994]).

40. Voir notamment Gebru, 2020; Prince et Schwarcz, 2020; Lambrecht et Tucker, 2019; Gillis et Spiess, 2019; Hu *et al.*, 2019; Williams, Brooks et Shmargad, 2018; Packin et Lev Aretz, 2018; Eubanks, 2018; Noble, 2018; Julia Angwin *et al.*, 2016; Barocas et Selbst, 2016.

41. Traduction libre de l'anglais vers le français.

sur des motifs protégés. Mentionnons un des exemples les plus connus en matière de discrimination algorithmique (non lié aux algorithmes de recommandation), à savoir le cas de l'algorithme COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) (Cofone, 2019 ; Packin et Lev Aretz, 2018 ; Wang, 2018). L'algorithme, visant à évaluer le risque de récidive des individus accusés, avait un préjugé négatif envers les personnes noires en les considérant comme davantage propices à commettre des crimes (Julia Angwin *et al.*, 2016). Cela avait lieu malgré le fait que l'algorithme n'avait pas accès à la « race » de l'individu. De façon similaire, les algorithmes de recommandation qui n'ont pas accès à des informations démographiques (qui sont souvent des motifs protégés au niveau de la discrimination) peuvent tout de même apprendre à faire des distinctions sur ces facteurs.

Notons également d'autres instances de biais. Une utilisatrice de Netflix critiquait, par exemple, le fait que dans le catalogue *Latino* il n'y avait que des *telenovelas* (Siles *et al.*, 2019, p. 511), reproduisant une vision stéréotypée du cinéma latino-américain. Au niveau du genre, il semble que les recommandations de Spotify montrent davantage des artistes masculins que des artistes féminins (Eriksson et Johansson, 2017 ; Ferraro *et al.*, 2021). Le fait que les recommandations soient genrées est également souligné par Werner (2020). Cette chercheuse énonce ainsi : « [l']organisation et la représentation des goûts musicaux et des groupes sociaux sur Spotify ne sont donc pas fortuites ou innocentes, mais renforcent les modèles de pouvoir » (Werner, 2020, p. 88)⁴². Ainsi, la question de la discrimination des algorithmes de recommandation touche non seulement les personnes utilisatrices, mais également les personnes créatrices de contenu.

CONCLUSION

Les algorithmes de recommandation sont de plus en plus présents dans l'univers numérique et leurs effets sont donc de plus en plus importants. En effet, ces algorithmes influencent grandement la façon dont les individus ont accès à des films, des chansons, des publicités pour une panoplie de marchandises, des nouvelles sur l'actualité, etc. Nous avons montré que, bien qu'ils ne déterminent pas automatiquement les goûts, les opinions et les comportements, leur influence sur la construction de l'identité des individus n'est pas négligeable. Le concept de gouvernementalité

42. Traduction libre de l'anglais vers le français.

algorithmique, avec son accent sur le fait que ces algorithmes imposent une norme évolutive individualisée délimitant l'espace des possibilités, nous a donc été utile. Nous avons expliqué que les algorithmes de recommandation menacent la liberté d'opinion et la liberté d'expression en ce qu'ils peuvent limiter le contenu culturel et médiatique auquel les individus ont accès et façonner leurs goûts et leurs opinions. Nous avons également montré que le droit à la vie privée est lui aussi menacé par ces algorithmes, autant en matière de protection des données personnelles qu'en matière de sphère d'autonomie sans interférence. Les algorithmes de recommandation, par le fait qu'ils effectuent des profils, donc des distinctions entre les individus, peuvent aussi parfois effectuer des distinctions sur des motifs protégés. Cette potentielle discrimination algorithmique représenterait une atteinte au droit à l'égalité. Bien que notre analyse se soit limitée à ces trois droits, il serait pertinent de poursuivre la réflexion et d'analyser les répercussions que les algorithmes de recommandation ont sur d'autres droits et libertés, tels que les droits culturels prévus dans la DUDH et le *Pacte international relatif aux droits économiques, sociaux et culturels*⁴³.

RÉFÉRENCES

Instruments juridiques

Traités internationaux

Pacte international sur les droits économiques, sociaux et culturels, 999 RTNU 3, 16 décembre 1966 (entrée en vigueur : 3 janvier 1976) [PIDESC].

Pacte international relatif aux droits civils et politiques, 999 RTNU 171, 19 décembre 1966 (entrée en vigueur : 23 mars 1976) [PIDCP].

Déclarations

Déclaration universelle des droits de l'homme, Rés AGA 217A (III), Doc off AG NU, 3^e sess, supp n° 13, Doc NU A/810 (1948) 71.

Droit dérivé

Observation générale n° 16 : Article 17 (Droit au respect de la vie privée), Doc off Comité des droits de l'homme 32^e sess, Doc NU HRI\GEN\1\Rev.1 (1988) [1994].

Observation générale n° 18 : Non-discrimination, Doc off Comité des droits de l'homme, 37^e sess, Doc NU HRI\GEN\1\Rev.1 (1989) [1994].

Observation générale n° 34 : Article 19 : Liberté d'opinion et liberté d'expression, Doc off Comité des droits de l'homme, 102^e sess, Doc NU CCPR/C/GC/34 (2011).

43. *Pacte international sur les droits économiques, sociaux et culturels*, 999 RTNU 3, 16 décembre 1966 (entrée en vigueur : 3 janvier 1976) [PIDESC].

Principes directeurs relatifs aux entreprises et aux droits de l'homme : mise en œuvre du cadre de référence «protéger, respecter et réparer» des Nations Unies, Doc off Conseil des droits de l'homme, 17^e sess, annexe, Doc NU A/HRC/17/31 (2011) 7.

Doctrine

Access Now, *Human Rights in the Age of Artificial Intelligence*, New York, 2018.

Adomavicius, Gediminas et Alexander Tuzhilin, « Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions », *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, n° 6, 2005, p. 734749.

Agre, Philip E. et Marc Rotenberg (dir.), *Technology and Privacy. The New Landscape*, MIT Press, 1998.

Aguirre Hernando, Clara, « Backstage to the Panama Papers : Big data analytics and collaborative journalism », *Polis LSE*, 2017.

Angwin, Julia, Surya Mattu, Lauren Kirchner et Jeff Larson, *Machine Bias*. ProPublica, 2016, <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>.

Art AI Gallery, « Art made by artificial intelligence : Art that captures the eye & the mind », n.d., <<https://www.artaigallery.com/>>.

Barocas, Solon, et Andrew D. Selbst, « Big data's disparate impact », *California Law Review*, vol. 104, n° 3, 2016, p. 671732.

Bayefsky, Anne F., « The principle of equality or non-discrimination in international law », dans Stephanie Farrior (dir.), *Equality and Non-Discrimination under International Law : Volume II*, Ashgate, 2015.

Beuscart, Jean-Samuel, Samuel Coavoux et Sisley Maillard, « Les algorithmes de recommandation musicale et l'autonomie de l'auditeur : Analyse des écoutes d'un panel d'utilisateurs de streaming », *Réseaux*, vol. 213, n° 1, 2019, p. 17-47.

Bezemek, Christoph, « The "filter bubble" and human rights », dans Bilyana Petkova et Tuomas Ojanen (dir.), *Fundamental Rights Protection Online*, Edward Elgar Publishing, 2020.

Bucher, Taina, *If... Then : Algorithmic Power and Politics*, New York, Oxford University Press, 2018.

Burke, Robin, Alexander Felfernig et Mehmet H. Göker, « Recommender systems : An overview », *AI Magazine*, vol. 32, n° 3, 2011, p. 1318.

Cabra, Mar, « How ICIJ cracked the Panama and the paradise papers », *Big Data Spain*, 2017 [conférence].

Calandrino, Joseph A., Ann Kilzer, Arvind Narayanan, Edward W. Felten et Vitaly Shmatikov, « "You might also like." Privacy risks of collaborative filtering », *2011 IEEE Symposium on Security and Privacy*, 2011.

Castagnos, Sylvain, Armelle Brun et Anne Boyer, « La diversité : entre besoin et méfiance dans les systèmes de recommandation », *Revue I3 – Information, interaction, intelligence*, Cepaduès, 2014.

Christie's, « Is artificial intelligence set to become art's next medium ? », *Christie's*, 2018, <<https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>>.

Cofone, Ignacio N., « Algorithmic discrimination is an information problem », *Hastings Law Journal*, vol. 70, n° 6, 2019, p. 13891444.

Cohen, Julie E, « What privacy is for », *Harvard Law Review*, vol. 126, n° 7, 2013, p. 19041932.

- Comité d'experts sur les intermédiaires d'Internet (MSI-NET), *Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données (en particulier les algorithmes) et éventuelles implications réglementaires*, Strasbourg, Conseil de l'Europe, 2017.
- Datta, Amit, Michael Carl Tschantz et Anupam Datta, « Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination », *Proceedings on Privacy Enhancing Technologies*, n° 1, 2015, p. 92112.
- de Zayas, Alfred et Áurea Roldán Martín, « Freedom of opinion and freedom of expression: Some reflections on General Comment n° 34 of the UN Human Rights Committee », *Netherlands International Law Review*, vol. 59, n° 3, 2012, p. 425454.
- Dubois, Lorette et Florence Gaullier, « Publicité ciblée en ligne, protection des données à caractère personnel et ePrivacy : un ménage à trois délicat », *LEGICOM*, vol. 59, n° 2, 2017, p. 69102.
- Eriksson, Maria C. et Anna Johansson, « Tracking gendered streams », *Culture Unbound*, vol. 9, n° 2, 2017, p. 163183.
- Eskens, Sarah, Natali Helberger et Judith Moeller, « Challenged by news personalisation: Five perspectives on the right to receive information », *Journal of Media Law*, vol. 9, n° 2, 2017, p. 259-284.
- Eskens, Sarah, « The personal information sphere: An integral approach to privacy and related information and communication rights », *Journal of the Association for Information Science and Technology*, vol. 71, n° 9, 2020, p. 11161128.
- Eubanks, Virginia, *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor*, New York, St. Martin's Press, 2018.
- Ferraro, Andres, Xavier Serra et Christine Bauer, « Break the Loop: Gender Imbalance in Music Recommenders », *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, New York, Association for Computing Machinery, 2021.
- Flaxman, Seth, Sharad Goel et Justin M. Rao, « Filter bubbles, echo chambers, and online news consumption », *Public Opinion Quarterly*, vol. 80, n° S1, 2016, p. 298320.
- Flores Echaiz, Lucia, « Artificial intelligence in Canada and equality: Algorithmic discrimination and section 15(1) of the charter », Nasma Ahmed, Suzie Dunn et Florian Martin-Bariteau (dir.), *Can't Compute*, Ottawa, University of Ottawa Research Chair in Technology and Society, 2023.
- Fredman, Sandra, *Discrimination Law*, 2^e éd., New York, Oxford University Press, 2011.
- Gebru, Timnit, « Race and gender », dans Markus Dubber, Frank Pasquale et Sunit Das (dir.), *Oxford Handbook on AI Ethics*, New York, Oxford University Press, 2020.
- Gillis, Talia B. et Jann L. Spiess, « Big Data and discrimination », *The University of Chicago Law Review*, vol. 86, n° 2, 2019, p. 459488.
- Gottron, Thomas et Felix Schwagereit, « The impact of the filter bubble – A simulation based framework for measuring personalisation macro effects in online communities », *arXiv:1612.06551 [cs]*, 2016.
- Groupe de travail « Article 29 » sur la protection des données, *Avis 2/2010 sur la publicité comportementale en ligne*, 00909/10/FR, WP 171, 2010.
- Hallinan, Blake et Ted Striphas, « Recommended for you: The Netflix prize and the production of algorithmic culture », *New Media & Society*, vol. 18, n° 1, 2016, p. 117137.

- Hu, Xianhong, Lucia Flores Echaiz, Bhanu Neupane, Prateek Sibal et Macarena Rivera Lam, *Steering AI and Advanced ICTs for Knowledge Societies: A Rights, Openness, Access, and Multi-stakeholder Perspective*, Paris, UNESCO, 2019.
- Huq, Proma, « Music to my ears : De-Blackboxing Spotify's recommendation engine », *CCTP-607: « Big Ideas »: AI to the Cloud*, 2019, <<https://blogs.commonsgorgetown.edu/cctp-607-spring2019/2019/05/06/music-to-my-ears-de-blackboxing-spotifys-recommendation-algorithm/>>.
- Johnson, Christopher C., « Logistic matrix factorization for implicit feedback data », *Advances in Neural Information Processing Systems*, vol. 27, n° 78, p. 1-9, 2014.
- Karakayali, Nedim, Burc Kostem et Idil Galip, « Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste », *Theory, Culture & Society*, vol. 35, n° 2, 2018, p. 324.
- Lambrech, Anja et Catherine Tucker, « Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads », *Management Science*, vol. 65, n° 7, 2019, p. 29662981.
- Lekakos, George et George M. Giaglis, « Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors », *Interacting with Computers*, vol. 18, n° 3, 2006, p. 410431.
- Lemelshtrich Latar, Noam, « The robot journalist in the age of social physics: The end of human journalism? », dans Gali Einav (dir.), *The New World of Transitioned Media*, Springer, 2015.
- Liu, Ping, Karthik Shivaram, Aron Culotta, Matthew A. Shapiro et Mustafa Bilgic, « The interaction between political typology and filter bubbles in news recommendation algorithms », *Proceedings of the Web Conference 2021*, 2021.
- Lury, Celia et Sophie Day, « Algorithmic personalization as a mode of individuation », *Theory, Culture & Society*, vol. 36, n° 2, 2019, p. 1737.
- Maillet, François, *Algorithmes de recommandation musicale*, Université de Montréal, 2009.
- McCarthy, John, « Sony uses AI to compose a Beatles-inspired song, the unforgettable "Daddy's Car" », *The Drum*, 2016, <<https://www.thedrum.com/news/2016/09/27/sony-uses-ai-compose-beatles-inspired-song-the-unforgettable-daddy-s-car>>.
- McEnnis, Daniel et Sally Jo Cunningham, « Sociology and music recommendation systems », *Proceedings of the 8th International Conference on Music Information Retrieval*, 2007.
- McMillan, John et Jeanne Snelling, « Equality: Old debates, new technologies », dans Roger Brownsword, Eloise Scotford et Karen Yeung (dir.), *The Oxford Handbook of Law, Regulation and Technology*, Oxford University Press, 2017.
- Ménard, Marc, « Systèmes de recommandation de biens culturels », *Les Cahiers du numérique*, vol. 10, n° 1, 2014, p. 6994.
- Morris, Jeremy Wade, Robert Prey et David B. Nieborg, « Engineering culture: Logics of optimization in music, games, and apps », *Review of Communication*, vol. 21, n° 2, 2021, p. 161175.
- Moses, Lucia, « The Washington Post's robot reporter has published 850 articles in the past year », *Digiday*, 2017, <<https://digiday.com/media/washington-posts-robot-reporter-published-500-articles-last-year/>>.
- Narayanan, Arvind et Vitaly Shmatikov, « How to break anonymity of the Netflix prize dataset », *ArXiv:Cs/0610105*, 2008.

- Nechushtai, Efrat, et Seth C. Lewis, « What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations », *Computers in Human Behavior*, vol. 90, 2019, p. 298307.
- Nell, Alice, « Recommended culture: Distinctions in cultural consumption and taste in a digitized and recommendation driven age », University of Utrecht, 2014.
- Noble, Safiya Umoja, *Algorithms of Oppression: Race, Gender and Power in the Digital Age*, New York, NYU Press, 2018.
- O'Dair, Marcus et Andrew Fry, « Beyond the black box in music streaming: The impact of recommendation systems upon artists », *Popular Communication*, vol. 18, n° 1, 2020, p. 6577.
- Packin, Nizan Geslevich, et Yafit Lev Aretz, « Learning algorithms and discrimination », dans Barfield Woodrow et Ugo Pagallo (dir.), *Research Handbook on the Law of Artificial Intelligence*, Cheltenham, Edward Elgar Publishing, 2018.
- Pariser, Eli, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, New York, Penguin, 2011.
- Park, Deuk Hee, Hyea Kyeong Kim, Il Young Choi et Jae Kyeong Kim, « A literature review and classification of recommender systems research », *Expert Systems with Applications*, vol. 39, n° 11, 2012, p. 1005910072.
- Pasquale, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.
- Prince, Anya et Daniel Schwarcz, « Proxy discrimination in the age of artificial intelligence and big data », *Iowa Law Review*, vol. 105, n° 3, 2020, p. 12571318.
- Rapport du Haut-Commissaire des Nations Unies aux droits de l'homme (Le droit à la vie privée à l'ère du numérique)*, Doc off AG NU, Conseil des droits de l'homme, 39^e sess, Doc NU A/HRC/39/29 (2018).
- Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression*, Doc off AG NU, 73^e sess, Doc NU A/73/348 (2018).
- Rapport du Rapporteur spécial sur le droit à la vie privée*, Doc off AG NU, Conseil des droits de l'homme, 40^e sess, Doc NU A/HRC/40/63 (2019).
- Rengel, Alexandra, « Privacy as an international human right and the right to obscurity in cyberspace », *Groningen Journal of International Law*, vol. 2, n° 2, 2014, p. 33-54.
- Roberge, Jonathan, Romauld Jamet et Andréanne Rousseau, *L'impact social des algorithmes de recommandation sur la curation des contenus musicaux francophones au Québec: une revue de littérature comparée*, Institut national de la recherche scientifique, 2018.
- Rodriguez, Marie-Emmanuelle, *La publicité ciblée: un défi pour la protection des données à caractère personnel. Analyse d'un cas pratique: le ciblage publicitaire opéré par Facebook au regard des exigences imposées par le Règlement général sur la protection des données*, Université catholique de Louvain, 2020.
- Rouvroy, Antoinette et Bernard Stiegler, « Le régime de vérité numérique. De la gouvernementalité algorithmique à un nouvel État de droit », *Socio. La nouvelle revue des sciences sociales*, n° 4, 2015, p. 113140.
- Rouvroy, Antoinette et Thomas Berns, « Le nouveau pouvoir statistique », *Multitudes*, vol. 40, n° 1, 2010, p. 88103.
- Rouvroy, Antoinette et Thomas Berns. « Gouvernementalité algorithmique et perspectives d'émancipation », *Réseaux*, vol. 177, 2013, p. 163196.

- Rouvroy, Antoinette, « Of Data and Men ». *Fundamental Rights and Freedoms in a World of Big Data*, Strasbourg, Conseil de l'Europe, 2016a.
- Rouvroy, Antoinette, « Des données sans personne : le fétichisme de la donnée à caractère personnel à l'épreuve de l'idéologie des Big Data » dans *Étude annuelle du Conseil d'État : le numérique et les droits fondamentaux*, Paris, La Documentation française, 2014.
- Rouvroy, Antoinette, « Homo juridicus est-il soluble dans les données ? », dans Elise Degrave, Cécile de Terwangne, Séverine Dusollier, Robert Queck (dir.), *Droit, normes et libertés dans le cybermonde : Liber amicorum Yves Poullet*, Bruxelles, Larcier, 2018.
- Rouvroy, Antoinette, « La gouvernementalité algorithmique : radicalisation et stratégie immunitaire du capitalisme et du néolibéralisme ? », *La Deleuziana*, n° 3, 2016b, p. 3036.
- Siles, Ignacio, Johan Espinoza-Rojas, Adrián Naranjo et María Fernanda Tristán, « The Mutual Domestication of Users and Algorithmic Recommendations on Netflix », *Communication, Culture and Critique*, vol. 12, n° 4, 2019, p. 499518.
- Singel, Ryan, « Netflix cancels recommendation contest after privacy lawsuit », *Wired*, 2019, <<https://www.wired.com/2010/03/netflix-cancels-contest/>>.
- Spotify, « 6 questions (and answers) with Tony Jebara, VP of machine learning », *Newsroom Spotify*, 2019, <<https://newsroom.spotify.com/2019-10-24/6-questions-and-answers-with-tony-jebara-vp-of-machine-learning/>>.
- Striphas, Ted, « Algorithmic culture », *European Journal of Cultural Studies*, vol. 18, n°s 4-5, 2015, p. 395-412.
- Varela Martinez, Daniela Renee, *The Netflix Experience, Re-shaping the Creative Process: Cultural Co-Production of Content*, Huddinge, Södertörn University, 2017.
- Varela, Daniela et Anne Kaun, « The Netflix experience: A user-focused approach to the Netflix recommendation algorithm », dans Theo Plothe et Amber M. Buck (dir.), *Netflix at the Nexus: Content, Practice, and Production in the Age of Streaming Television*, New York, Peter Lang Publishing Group, 2019.
- Wang, Jackie, *Carceral Capitalism*, Cambridge, MIT Press, 2018.
- Werner, Ann, « Organizing music, organizing gender: Algorithmic culture and Spotify recommendations », *Popular Communication*, vol. 18, n° 1, 2020, p. 7890.
- Williams, Betsy Anne, Catherine F. Brooks et Yotam Shmargad, « How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications », *Journal of Information Policy*, vol. 8, 2018, p. 78115.
- Wilson-Barnao, Caroline, « How algorithmic cultural recommendation influence the marketing of cultural collections », *Consumption Markets & Culture*, vol. 20, n° 6, 2017, p. 559574.

INTÉGRER LA DIVERSITÉ CULTURELLE DANS LES INSTRUMENTS RELATIFS À L'ENCADREMENT DE L'IA

VERS UNE TECHNOLOGIE *CULTURELLEMENT* RESPONSABLE ?

Iris Richer¹ et Clémence Varin²

Ces dernières années se distinguent par l'émergence d'un recours exponentiel à l'intelligence artificielle (IA) dans l'ensemble des secteurs culturels, que ce soit dans le domaine des industries culturelles et créatives ou du patrimoine. L'IA offre, par exemple, des outils novateurs d'accompagnement pour les artistes et auteurs, participant ainsi à la réinvention de certains processus créatifs. Dans le domaine du patrimoine, elle permet également d'apporter des solutions à plusieurs défis auxquels le secteur est confronté. L'IA peut ainsi être utilisée pour le contrôle du trafic illicite de biens culturels ou encore pour surveiller les éventuelles dégradations de biens culturels protégés causées par les conflits armés ou les changements climatiques (COMEST, 2019). La traduction automatique, pour sa part, est à même de constituer un formidable outil de modélisation de langues rares ou en voie de disparition, et de devenir un puissant vecteur de promotion de la diversité et de protection des langues, notamment autochtones (COMEST, 2019). Enfin, la manifestation la plus visible de l'utilisation

1. Doctorante en droit à l'Université Laval.

2. Doctorante en droit en cotutelle entre l'Université Laval et l'Université de Rennes.

de l'IA (et des algorithmes) dans le secteur culturel provient des plateformes numériques, à l'instar de Spotify ou Netflix. Celles-ci font appel à ces technologies pour recommander des contenus à leurs utilisateurs et, dans certains cas, pour les créer afin qu'ils correspondent aux préférences de ces derniers. Cependant, si ces systèmes offrent d'incroyables potentiels, ils font également émerger un certain nombre de défis en matière de préservation de la diversité culturelle, sur lesquels cette contribution se concentrera plus spécifiquement.

La diversité culturelle peut être définie comme « l'originalité et la pluralité des identités qui caractérisent les groupes et les sociétés composant l'humanité »³. Parmi les éléments qui la constituent figurent « l'ensemble des traits distinctifs spirituels et matériels, intellectuels et affectifs qui caractérisent une société ou un groupe social », ce qui englobe notamment les arts et les lettres, les systèmes de valeurs, ou encore les traditions et les croyances⁴. Ainsi, la diversité culturelle s'exprime à plusieurs niveaux, à la fois « *between individual states, regions, communities, individuals, but also within states, regions and communities* » (Donders, 2016, p. 24).

Dans une société aux prises avec un développement rapide et sans précédent des technologies numériques, la préservation de la diversité culturelle constitue désormais une préoccupation majeure. Or, elle est souvent absente des réflexions portées sur les nouvelles technologies, les conditions de leur utilisation et leur encadrement. La diversité culturelle est pourtant indispensable à la préservation d'une coexistence pacifique et harmonieuse entre les diverses cultures et pour favoriser un dialogue interculturel, essentiel à l'atteinte d'une véritable cohésion sociale et un meilleur vivre-ensemble.

En adoptant une perspective juridique, ce chapitre vise à démontrer l'importance d'inclure les préoccupations culturelles dans les instruments visant à encadrer l'IA afin qu'elle soit à l'image de la société, c'est-à-dire diversifiée et respectueuse de l'ensemble des cultures. Sans prétendre dresser un portrait exhaustif, la première partie de cette contribution est consacrée à la présentation de quelques-uns des risques⁵ que fait peser

3. *Déclaration universelle de l'UNESCO sur la diversité culturelle*, Rés CG, Doc off UNESCO, 31e sess, Doc UNESCO CLT-2002/WS/9 (2001), art. 1.

4. *Ibid.*, préambule.

5. Dans cet article, le terme « risque » est défini comme la « [p]ossibilité, probabilité d'un fait, d'un événement considéré comme un mal ou un dommage », tandis que le terme « menace » est entendu dans ce contexte comme un « [s]igne, indice qui laisse prévoir quelque chose ».

l'IA sur la diversité culturelle (1.1) et à l'étude de plusieurs instruments juridiques internationaux relatifs à la diversité culturelle et leur pertinence au regard des dangers de l'IA (1.2). Il s'agira ensuite d'analyser certains cadres et instruments adoptés pour réguler ces systèmes afin d'identifier d'éventuelles lacunes relatives à la protection de la diversité culturelle (2.1). Enfin, quelques recommandations seront formulées pour encourager le développement d'une IA *culturellement* responsable (2.2).

1. LA DIVERSITÉ CULTURELLE ET L'IA

La protection de la diversité culturelle est la célébration de notre identité culturelle, individuelle et collective, et la préservation de nos différences comme socle de notre humanité. À cet égard, toute avancée technologique qui bouleverse notre société doit être analysée au regard des impacts qu'elle provoque sur la préservation de cette diversité. Ainsi, une réglementation de l'IA ne peut être considérée sans qu'une réflexion approfondie soit menée sur les interactions entre les risques provoqués (1.1) et les instruments de droit international adoptés en matière de diversité culturelle (1.2).

1.1 Un état des lieux des risques de l'IA pour la diversité culturelle

Le développement de systèmes d'IA ne prenant pas en compte la diversité culturelle peut être à l'origine de multiples atteintes à cette dernière. Dans ce contexte, nous nous focalisons principalement sur deux catégories de risques, ceux qui sont effectivement liés au développement des systèmes d'IA et ceux qui résultent plus généralement du processus de mondialisation.

En effet, la mondialisation a créé des opportunités de dialogue et de coexistence renouvelées entre les cultures. Or, elle pose également des menaces sérieuses à la diversité culturelle, notamment en raison de l'évolution rapide des technologies numériques⁶. Une des premières menaces est celle de l'*homogénéisation culturelle*, qui se traduit par une uniformisation progressive des pratiques et des identités culturelles, en particulier sous l'influence des cultures dominantes. L'amplification de ce phénomène pose de nombreuses interrogations quant à la capacité des États à protéger les

de dangereux, de nuisible» (*Larousse*). Le risque se distingue donc par sa spécificité, alors que la menace s'inscrit quant à elle dans un cadre plus global (par exemple, une *menace* à la préservation de la diversité culturelle).

6. *Déclaration universelle de l'UNESCO sur la diversité culturelle*, *supra* note 3, préambule.

cultures présentes sur leur territoire par des politiques adaptées (Bernier et Atkinson, 2000, p. 8 ; Bernier et Ruiz Fabri, 2002, p. 23-24). Les conséquences de ce processus sur la diversité culturelle sont multiples et peuvent se traduire par un appauvrissement progressif des cultures, notamment les plus fragiles comme celles des minorités ou des peuples autochtones, et conduire à la disparition de certaines langues et pratiques culturelles (Bernier et Atkinson, 2000 ; Lewis *et al.*, 2020).

Face à ces défis, le développement progressif de l’IA paraît avoir le potentiel de fortement aggraver la portée de tels risques et ce, en raison des caractéristiques constitutives de ces systèmes. En effet, la reproduction et la propagation de biais préexistants par l’IA sont très largement documentées (Institut Montaigne, 2020 ; Raso *et al.*, 2018). En lien avec la diversité culturelle plus spécifiquement, nous pouvons distinguer notamment des risques liés à la dimension culturelle des biais algorithmiques et des risques liés à l’utilisation de données biaisées par ces systèmes d’IA.

Premièrement, la dimension culturelle des biais est souvent peu considérée, alors qu’elle participe pourtant activement au processus d’homogénéisation culturelle. En effet, les systèmes d’IA développés dans un contexte culturel donné peuvent véhiculer des valeurs culturelles propres à un pays ou une région du monde, et ne devraient donc pas être déployés uniformément sur l’ensemble des territoires sous peine de nuire à la préservation de la diversité culturelle. Pour cette raison, la prise en compte de « la dimension culturelle des biais algorithmiques ne [devrait] pas être ignorée » (Institut Montaigne, 2020, p. 34).

D’autre part, les données qui sont utilisées par ces systèmes d’IA peuvent comporter des biais qui s’expriment, par exemple, à l’encontre d’un groupe ou d’une communauté culturelle. Ainsi, l’utilisation de données biaisées concernant certains groupes marginalisés comme les peuples autochtones est notamment susceptible d’entraîner une « reprodu[ction] des représentations coloniales » de ces peuples (Gentelet et Bahary-Dionne, 2021, p. 193). Dans cette perspective, la Commission mondiale d’éthique des connaissances scientifiques et des technologies (COMEST) alerte sur le fait que les systèmes d’IA peuvent conduire à des situations de discrimination et « inculquer et exacerber des préjugés, entraînant de l’inégalité et de l’exclusion et menaçant la diversité culturelle » (COMEST, 2019, p. 9).

Par ailleurs, un système d’IA pourrait être utilisé pour identifier certains individus sur le fondement de leur identité culturelle. La censure

des contenus (Kulesz, 2018, p. 11) devenant dès lors un moyen de limiter la participation à la vie culturelle de certains individus et groupes culturels.

Outre ces risques portés à la préservation de la diversité culturelle, il est possible d'entrevoir également des atteintes portées plus précisément à la *diversité des expressions culturelles*⁷ par ces systèmes d'IA.

En effet, le recours aux technologies numériques de plus en plus performantes et à l'IA a fait émerger diverses menaces, notamment en matière d'accès à une diversité d'expressions culturelles (Kulesz, 2018). L'entrée fracassante des plateformes numériques dans le secteur de la culture a permis un accès sans précédent à une diversité de contenus culturels. Or, pour les recommander à leurs utilisateurs, celles-ci ont pour la plupart recours aux algorithmes et dans certains cas à l'IA (Hunt et McKelvey, 2019). De nombreuses études ont démontré les limites de ces technologies, leur opacité et leur effet négatif sur la diversité des expressions culturelles (Kulesz et Dutoit, 2020, p. 40-41 ; Bensamoun et Farchy, 2020). Si ces algorithmes occupent une place centrale pour faciliter l'accès à tel ou tel contenu du catalogue de ces plateformes, ils ne leur accordent pas la même visibilité – ou « découvrabilité » – pour une multitude de raisons, dont des choix de conception, éditoriaux, commerciaux ou encore des données de référencement incomplètes. Le manque de découvrabilité des contenus francophones en est un exemple, tout comme ceux des minorités ou des peuples autochtones (Guèvremont *et al.*, 2019, p. 154-158).

Une telle situation limite donc l'accès des individus à certaines expressions culturelles, et a des répercussions plus larges sur le secteur culturel. En effet, une mauvaise visibilité (ou découvrabilité) de ces contenus a une influence sur la rémunération des artistes et des professionnels de la culture, et donc sur la « viabilité économique du travail créatif » (Hunt et McKelvey, 2019, p. 4). Une autre problématique réside dans le fait que ces technologies sont développées, pour la plupart, par des acteurs

7. La *Convention sur la protection et la promotion de la diversité des expressions culturelles* de l'UNESCO (2005), qui traite d'un aspect particulier de la diversité culturelle à savoir la diversité des *expressions culturelles*, précise que la diversité culturelle renvoie à la « multiplicité des formes par lesquelles les cultures des groupes et des sociétés trouvent leur expression » et que « [c]es expressions se transmettent au sein des groupes et des sociétés et entre eux » (art 4.1). Les « expressions culturelles » sont définies comme « les expressions qui résultent de la créativité des individus, des groupes et des sociétés, et qui ont un contenu culturel » (art 4.3), et le contenu culturel réfère pour sa part « au sens symbolique, à la dimension artistique et aux valeurs culturelles qui ont pour origine ou expriment des identités culturelles » (art 4.2). *Convention sur la protection et la promotion de la diversité des expressions culturelles*, 2440 R.T.N.U. 311, 20 octobre 2005 (entrée en vigueur : 18 mars 2007) [Convention de 2005].

privés. Or, la « logique commerciale » de ces entreprises « risque d’entraîner une concentration accrue de l’offre, des données et des revenus, [et provoquer] à long terme un appauvrissement des expressions culturelles » (Kulesz, 2018, p. 14). Dans une telle situation, ces expressions sont susceptibles de se retrouver amputées « d’identité et de signification propres » (Kulesz, 2018, p. 8).

Le risque d’uniformisation est d’autant plus présent que ces systèmes accentuent la similarité des biens culturels consommés et celle des utilisateurs, ce que l’on peut qualifier d’« homophilie »⁸ (Hoffman *et al.* dans Burri, 2019, p. 5). Parmi les autres effets, on peut citer les risques de manipulation des utilisateurs (Latzer *et al.* dans Burri, 2019, p. 9), de diminution de la culture publique (Hunt et McKelvey, 2019, p. 8), l’impact sur la liberté de choix des individus (Burri, 2019, p. 5), ainsi que la création d’une « bulle parfaite » autour des utilisateurs (Kulesz, 2018, p. 5).

Ainsi, des systèmes d’IA inadaptés ou biaisés peuvent avoir des répercussions considérables sur la diversité culturelle. Particulièrement insidieux, de tels impacts ne seront visibles qu’à long terme, ce qui conduit à un manque actuel de considération à l’égard de ces risques⁹, tout comme d’une volonté d’y répondre. Or, un cadre international existe pour mitiger ce genre de risques et éviter que des atteintes irrémédiables soient portées à la préservation de la diversité culturelle.

1.2 Les instruments de droit international relatifs à la diversité culturelle au regard des défis posés par l’IA

Les États ont adopté des dispositions internationales qui devraient être prises en compte au moment de l’élaboration ou de l’utilisation de systèmes d’IA, y compris pour la préservation de la diversité culturelle. En effet, un certain nombre de risques identifiés font écho à des préoccupations qui ont, d’ores et déjà, été intégrées au sein de divers instruments juridiques internationaux des droits de la personne (1.2.1) et du droit international de la culture (1.2.2).

8. Ce concept désigne la tendance d’un individu à s’affilier à ses semblables.

9. Ce manque de considération peut s’expliquer à la fois par le manque de données collectées et d’études menées sur ce sujet, mais aussi plus largement par une difficile prise de conscience des conséquences possibles de ces systèmes sur la diversité culturelle.

1.2.1 *Le droit international des droits de la personne : un socle juridique aux origines de la préservation de la diversité culturelle*

Le droit international des droits de la personne affirme que « [t]ous les peuples ont le droit de disposer d'eux-mêmes. En vertu de ce droit, ils [...] assurent librement leur développement économique, social et culturel »¹⁰. Cette disposition fondamentale reconnaît à la fois l'égalité de tous les peuples et leur droit d'être différents, notamment sur le fondement de leur culture. Ainsi, bien que la préservation de la diversité culturelle fasse rarement l'objet d'une mention explicite, elle est pourtant intégrée dans plusieurs dispositions des instruments internationaux adoptés. Une telle analyse est d'autant plus pertinente pour régler l'IA que le cadre juridique des droits de la personne fait l'objet d'un large consensus à l'international (Donahoe et MacDuffee Metzger, 2019, p. 116).

La diversité culturelle apparaît tout d'abord dans le champ d'application du droit à l'égalité, entendu comme le droit de ne pas être discriminé sur des considérations fondées « sur la race, la couleur, le sexe, la langue, la religion, l'opinion politique ou toute autre opinion [et] l'origine nationale [...] ». Ce dernier a été reconnu comme un droit fondamental inscrit à la fois dans le *Pacte international relatif aux droits civils et politiques* (PIDCP)¹¹ et le *Pacte international relatif aux droits économiques, sociaux et culturels* (PIDESC)¹². Bien qu'elle ne soit pas nommée explicitement, Donders souligne que la préservation de la diversité culturelle doit être comprise comme étant incluse dans le champ d'application du droit à l'égalité (Donders, 2016, p. 25-26) puisque le respect de ce droit conditionne le respect de l'ensemble des droits humains.

Parmi ces droits humains figurent des droits qui entretiennent une relation privilégiée avec la diversité culturelle, à savoir les droits culturels. La *Déclaration universelle des droits de l'homme* affirme en effet que « [t]oute personne a le droit de prendre part librement à la vie culturelle de la communauté [et] de jouir des arts »¹³, tout comme le PIDESC qui érige

10. *Pacte international relatif aux droits civils et politiques*, 999 RTNU 171, 19 décembre 1966 (entrée en vigueur : 23 mars 1976) [PIDCP] ; *Pacte international sur les droits économiques, sociaux et culturels*, 999 RTNU 3, 16 décembre 1966 (entrée en vigueur : 3 janvier 1976) [PIDESC], art. 1.1 (nos italiques).

11. *Pacte international relatif aux droits civils et politiques*, *ibid.*, art. 2.1.

12. *Pacte international sur les droits économiques, sociaux et culturels*, *ibid.*, art. 2.2.

13. *Déclaration universelle des droits de l'homme*, Rés AG 217A (III), Doc off AGNU, 3^e sess, supp n° 13, Doc NU A/810 (1948) [DUDH], art. 27.

le droit de participer à la vie culturelle comme un droit fondamental¹⁴. Ainsi, bien qu'il n'y ait pas de véritable mention de la diversité culturelle, il est communément admis que la protection et la mise en œuvre de ces droits ont pour effet de la protéger (Donders, 2016, p. 26; Blake, 2014, p. 20; Déclaration de Fribourg, 2002, paragr. 4); ils sont donc interreliés et « se renforcent mutuellement » (Shaheed, 2010, p. 11). Par ailleurs, le droit à l'identité culturelle a progressivement été reconnu comme une des composantes du droit de participer à la vie culturelle (Romainville, 2016, p. 237), et se définit comme le droit « de choisir sa propre identité, de s'identifier ou non à une ou plusieurs communautés données ou de modifier ce choix »¹⁵. Là encore, plusieurs instruments reconnaissent que la diversité culturelle et l'identité culturelle « sont indissociables »¹⁶ et qu'elles doivent être mutuellement garanties à l'ensemble des individus. Enfin, rappelons que le droit de participer à la vie culturelle suppose le droit à la liberté d'expression, ce qui inclut le droit à la liberté artistique¹⁷ et le droit d'accéder à la culture « grâce à n'importe quel moyen technique d'information ou de communication »¹⁸.

Le développement des systèmes d'IA peut affecter l'ensemble des droits humains (Raso *et al.*, 2018, p. 4), mais les atteintes aux droits culturels sont généralement méconnues (Guèvremont *et al.*, 2019, p. 24). De même, bien que le droit à l'égalité et à la non-discrimination soit fortement présent dans le domaine de l'IA, la préservation de la diversité culturelle n'est généralement pas considérée comme un élément constitutif de ce droit. Il est pourtant reconnu que les instruments du droit international des droits de la personne s'appliquent également dans l'environnement numérique (Bennoune, 2019, p. 21). Soulignons, par ailleurs, que de telles dispositions, qui doivent être appliquées par les États signataires, incluent également la responsabilité de veiller à ce que ces droits soient respectés par les acteurs privés (Raso *et al.*, 2018, p. 12-13, Guèvremont *et al.*, 2021).

14. *Pacte international sur les droits économiques, sociaux et culturels*, *supra* note 10, art. 15.

15. *Observation générale n° 21 : Droit de chacun de participer à la vie culturelle* (Article 15, paragraphe 1.a), Doc off Comité des droits économiques, sociaux et culturels, 43^e sess, Doc NU E/C.12/GC/21 (2009) au paragr. 15.a).

16. *Déclaration de Mexico sur les politiques culturelles*, adoptée lors de la Conférence mondiale sur les politiques culturelles du 26 juillet au 6 août 1982, UNESCO, Mexico, Mexique (1982) [Déclaration de Mexico] au paragr. 5.

17. *Pacte international relatif aux droits civils et politiques*, *supra* note 10, art. 19.

18. *Observation générale n° 21 : Droit de chacun de participer à la vie culturelle*, *supra* note 15, paragr. 15.b).

Enfin, d'autres instruments internationaux relatifs à la diversité culturelle ont été adoptés à l'échelle internationale, et ces derniers doivent également être pris en compte afin de réglementer efficacement les impacts de l'IA sur la diversité culturelle.

1.2.2 *La contribution de l'UNESCO à la protection de la diversité culturelle à l'échelle internationale*

En matière de diversité culturelle, l'UNESCO est sans doute l'institution de prédilection en raison de ses nombreux travaux sur le sujet, notamment l'adoption de la *Déclaration universelle sur la diversité culturelle*¹⁹ et de la *Convention sur la protection et la promotion de la diversité des expressions culturelles* (Convention de 2005)²⁰.

Adoptée à l'unanimité, la *Déclaration universelle sur la diversité culturelle* érige la diversité culturelle au rang de « patrimoine commun de l'humanité ». Elle y est vue comme aussi nécessaire que « la biodiversité dans l'ordre du vivant », et comme devant « être reconnue et affirmée au bénéfice des générations présentes et des générations futures²¹ ». Sa défense est considérée comme « inséparable du respect de la dignité de la personne humaine »²², ce qui implique le respect des droits humains et la pleine réalisation des droits culturels. Dès son adoption, les États s'inquiètent des « mutations économiques et technologiques » et reconnaissent qu'il est nécessaire d'accorder une attention particulière à « la diversité de l'offre créatrice [...] ainsi qu'à la spécificité des biens et services culturels qui [...] ne doivent pas être considérés comme des marchandises ou des biens de consommation comme les autres »²³.

Si cette Déclaration n'est pas contraignante, elle ouvre la voie à l'adoption de la *Convention sur la protection et la promotion de la diversité des expressions culturelles*²⁴. En tant que traité, cette Convention réaffirme des droits et crée des obligations pour les 151 États qui en sont parties (ainsi que l'Union européenne). Ainsi, elle reconnaît la double nature – économique et culturelle – des biens et services culturels et le droit souverain des États

19. *Déclaration universelle de l'UNESCO sur la diversité culturelle*, supra note 3.

20. *Convention sur la protection et la promotion de la diversité des expressions culturelles*, supra note 7.

21. *Déclaration universelle de l'UNESCO sur la diversité culturelle*, supra note 3, art. 1.

22. *Ibid.*, art. 4.

23. *Ibid.*, art. 8 (nos italiques).

24. Il convient de préciser que cette Déclaration est accompagnée de lignes essentielles d'un Plan d'action pour sa mise en œuvre, lesquelles font expressément référence à « l'opportunité d'un instrument juridique international sur la diversité culturelle » au paragr. 1.

d'adopter des politiques et des mesures pour protéger et promouvoir la diversité des expressions culturelles sur leur territoire.

La Convention est particulièrement pertinente puisque les expressions culturelles visées font partie de celles qui sont directement impactées par les systèmes d'IA²⁵. Bien que ce traité soit technologiquement neutre (il s'applique quelles que soient les technologies utilisées), les Parties à la Convention ont adopté des *Directives opérationnelles pour la mise en œuvre de la Convention dans l'environnement numérique* afin de répondre adéquatement aux défis liés aux technologies numériques aux différents stades de la chaîne de valeur culturelle (création, production, distribution/diffusion et accès)²⁶. Ces directives visent à réaffirmer le droit souverain des Parties de protéger et promouvoir la diversité des expressions culturelles dans l'environnement numérique²⁷ et à y promouvoir un accès équitable aux biens et services culturels²⁸. Plusieurs dispositions ont aussi pour objectif de garantir un accès diversifié aux expressions culturelles dans cet environnement. D'une part, à l'étape de la distribution/diffusion, elles précisent que les mesures mises en place doivent « encourager la diversité des médias numériques, y compris la multiplicité des distributeurs numériques de biens et services culturels et des acteurs du numérique (plateformes en ligne, [...]), tout en garantissant la visibilité et la découvrabilité des contenus culturels nationaux et locaux »²⁹. D'autre part, à l'étape de l'accès, les directives reconnaissent l'importance « d'assurer l'accès libre et pérenne aux diverses expressions culturelles, et d'accroître la participation à la vie culturelle dans l'environnement numérique »³⁰, en mettant en place des mesures visant à « instaurer une plus grande transparence et une meilleure lisibilité des modes d'indexation et de référencement des contenus », et ce « afin que les mécanismes numériques (algorithmes de recommandation) qui déterminent les contenus disponibles pour les utilisateurs offrent un large éventail d'expressions culturelles diverses »³¹.

25. Non seulement les expressions culturelles sont recommandées par des algorithmes, mais elles peuvent également être créées par des IA et servir à les entraîner (Kulesz, 2018).

26. *Directives opérationnelles sur la mise en œuvre de la Convention dans l'environnement numérique*, approuvées par la Conférence des Parties à la Convention de 2005 sur la protection et la promotion de la diversité des expressions culturelles lors de sa sixième session, Paris (2017). Ces directives ont été complétées par une Feuille de route proposant aux Parties diverses activités afin de les mettre en œuvre.

27. *Ibid.*, au paragr. 8.4.

28. *Ibid.*, au paragr. 8.6.

29. *Ibid.*, au paragr. 16.1.

30. *Ibid.*, au paragr. 17.

31. *Ibid.*, au paragr. 17.1.

Ainsi, en ratifiant la Convention, les Parties se sont engagées à protéger et à promouvoir la diversité des expressions culturelles sur leur territoire et à l'échelle internationale, mais également dans l'environnement numérique. Ce traité, ainsi que la *Déclaration universelle sur la diversité culturelle*, dans une moindre mesure, justifie donc de veiller à ce que les algorithmes de recommandation, et plus largement les systèmes d'IA utilisés dans le secteur de la culture, garantissent l'accès à une diversité d'expressions culturelles³². Ces instruments supposent donc la prise en compte, par les États, de telles considérations dans les instruments développés pour encadrer l'IA.

Les instruments de droit international analysés constituent un cadre juridique qui doit servir de fondement pour réglementer l'IA dans la mesure où ils s'appliquent également dans l'environnement numérique. Face à l'adoption de règles peu contraignantes dans ce domaine, ces instruments conservent toute leur légitimité et leur pertinence. Cependant, il apparaît indispensable d'adopter, en parallèle, des instruments qui seront dédiés spécifiquement à la réglementation de l'IA et qui permettront de mettre en œuvre plus efficacement les dispositions des instruments juridiques susmentionnés (Ben-Israel *et al.*, 2020, p. 91) et ainsi répondre aux nouveaux défis liés à l'IA (Conseil de l'Europe, 2020, p. 22). En effet, les instruments juridiques internationaux sont porteurs de limites liées au caractère général de leurs dispositions, qui peuvent être difficiles à mettre en œuvre face aux spécificités de cette technologie (Conseil de l'Europe, 2020, p. 23, 25 ; Langlois et Régis, 2021, p. 159).

2. LA PRISE EN COMPTE DE LA DIVERSITÉ CULTURELLE DANS LES INSTRUMENTS RELATIFS À L'IA

La place croissante de l'IA dans nos sociétés a poussé la communauté internationale – entreprises privées, société civile, États – à entreprendre des initiatives visant à réglementer ses usages. Les systèmes d'IA requièrent un encadrement spécifique qui se doit de respecter et d'appliquer le droit international, notamment en raison de leurs impacts actuels et potentiels sur des droits protégés. Cette deuxième partie vise à comparer une sélection de cadres et d'instruments internationaux développés dans

32. En ratifiant la Convention, les Parties se sont également engagées à promouvoir les objectifs et principes du traité dans d'autres enceintes internationales. *Convention sur la protection et la promotion de la diversité des expressions culturelles*, *supra* note 7, art. 21.

le domaine de l’IA et à les analyser au regard de l’ensemble des instruments juridiques applicables en matière de diversité culturelle³³. Une telle démarche, bien qu’elle soit embryonnaire, vise à déterminer si ces engagements en matière de diversité culturelle sont adéquatement pris en considération dans les cadres et les instruments relatifs à l’IA (2.1), et à formuler quelques recommandations dans cette perspective (2.2).

2.1 Un état des lieux de la prise en compte de la diversité culturelle dans les cadres éthiques et instruments juridiques relatifs à l’IA

La *Déclaration universelle sur la diversité culturelle* affirme que « [l]a défense de la diversité culturelle est un impératif éthique, inséparable du respect de la dignité [...] humaine³⁴ ». On peut donc s’interroger sur la prise en compte de la diversité culturelle en tant que principe éthique dans les instruments visant à encadrer l’IA. Notre analyse portera sur les cadres éthiques développés par une pluralité d’acteurs, souvent non étatiques (2.1.1), puis sur la *Recommandation sur l’éthique de l’intelligence artificielle* de l’UNESCO qui constitue très certainement l’instrument le plus abouti en la matière actuellement (2.1.2).

2.1.1 La diversité culturelle dans les cadres éthiques de l’IA : une inclusion limitée

Ces dernières années, l’adoption de cadres éthiques s’est imposée « *as one component of developing AI governance* » (Langlois et Régis, 2021, p. 151). Les avantages de tels cadres sont non négligeables puisqu’ils ont constitué un premier pas vers une réflexion approfondie sur les impacts liés à ces systèmes d’IA (Langlois et Régis, 2021, p. 156), y compris les risques (Daelman, 2021, p. 147). Ils ont permis d’identifier des points de convergence et de consensus entre les acteurs qui les ont adoptés, lesquels sont indispensables pour envisager l’adoption d’instruments plus contraignants, tout particulièrement à l’échelle internationale (Langlois et Régis, 2021, p. 163).

Malgré ces éléments positifs, l’analyse de quelques cadres éthiques relatifs à l’IA permet de déceler des imprécisions, manquements ou incohérences eu égard à la notion de « diversité culturelle », ce qui limite

33. Cette analyse se concentre uniquement sur les instruments analysés dans la première partie de cette contribution, c’est-à-dire ceux émanant à la fois du droit international des droits de la personne et du droit international de la culture.

34. *Déclaration universelle de l’UNESCO sur la diversité culturelle*, *supra* note 3, art. 4.

l'atteinte des objectifs visés par ces derniers. Tout d'abord, l'inclusion de considérations relatives à la « diversité culturelle » n'est pas systématique, malgré les risques qui pèsent sur celle-ci. Dans les cas où elle est effectivement mentionnée, il est permis de douter du sens qui lui est donné tant les mentions sont furtives (Kulesz, 2018, p. 12). Dans la plupart des cas, aucune définition n'est fournie, ce qui ne permet pas de confirmer avec certitude qu'il s'agit de la « diversité culturelle » au sens du droit international. C'est le cas des *Principes d'Asilomar* qui énoncent que les systèmes d'IA doivent respecter les valeurs humaines et de ce fait « should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity »³⁵. Si la présence de la « diversité culturelle » mérite d'être soulignée, la succession de termes interroge sur leur réelle signification, d'autant qu'il existe rarement d'autres éléments pouvant être rapprochés de ce concept ailleurs dans le texte. Nous constatons également que le terme « diversité » ne renvoie pas toujours à la diversité culturelle au sens où nous l'entendons, et peut se limiter uniquement à celle des utilisateurs et des équipes de conception de ces systèmes. C'est le cas des *Lignes directrices en matière d'éthique pour une IA digne de confiance* du Groupe d'experts de haut niveau sur l'intelligence artificielle constitué par la Commission européenne³⁶. Celles-ci énoncent en effet que, « [p]our parvenir à une IA digne de confiance, il est nécessaire de favoriser l'inclusion et la diversité tout au long du cycle de vie du système d'IA [ce qui] implique également de veiller à l'égalité d'accès au moyen de processus conçus de manière inclusive, ainsi qu'à l'égalité de traitement »³⁷. Elles précisent par ailleurs qu'« il est essentiel que les équipes qui conçoivent, mettent au point, testent, entretiennent, déploient et/ou achètent ces systèmes reflètent la diversité des utilisateurs et de la société en général », afin de représenter « différents points de vue, besoins et objectifs ». Elles encouragent donc la constitution d'équipes diversifiées composées d'individus de tous les genres et de tous les âges et provenant de cultures différentes³⁸.

35. *Asilomar AI Principles*, Future of Life Institute, 2017 au principe 11 (nos italiques).

36. *Lignes directrices en matière d'éthique pour une IA digne de confiance*, Commission européenne, Groupe d'experts de haut niveau sur l'intelligence artificielle (GEHN IA), 8 avril 2019.

37. *Ibid.*, au paragr. 79. Il convient de préciser que cette volonté de diversité au sein des équipes de conception vise également à atténuer d'autres risques précédemment identifiés eu égard à la diversité culturelle, tels que les biais et préjugés que peuvent reproduire les systèmes d'IA.

38. *Ibid.*, au paragr. 111.

À l'inverse, la *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*³⁹ consacre un principe entier à « l'inclusion de la diversité » et précise que « [l]e développement et l'utilisation de SIA [systèmes d'intelligence artificielle] doivent être compatibles avec le maintien de la *diversité* sociale et *culturelle* et ne doivent pas restreindre l'éventail des choix de vie et des expériences personnelles »⁴⁰. Bien qu'on puisse regretter l'absence de mention des textes internationaux, la déclaration précise néanmoins que « [l]e développement et l'utilisation de SIA ne devraient pas conduire à une uniformisation de la société par la normalisation des comportements et des opinions, [et] prendre en considération les multiples expressions des diversités sociales et culturelles, et cela dès la conception des algorithmes ». On constate donc dans ce cas que le terme de « diversité » se rapproche de la diversité culturelle telle qu'elle est entendue dans ce chapitre.

La Déclaration de Toronto intitulée *Protéger le droit à l'égalité et à la non-discrimination dans les systèmes reposant sur l'apprentissage automatique* (2018), qui s'inscrit dans une perspective de droits humains, offre également un angle intéressant puisqu'elle explicite les préoccupations culturelles liées au droit à l'égalité. Outre le fait qu'elle souligne que ces systèmes doivent promouvoir la diversité⁴¹, elle reconnaît que les systèmes algorithmiques peuvent avoir des répercussions sur le droit des individus de participer à la vie culturelle, et donc les droits culturels⁴². La mention claire de ces enjeux mérite d'être soulignée en raison des insuffisances généralement identifiées. De manière générale, on observe qu'un certain nombre de cadres éthiques mentionnent que les droits de la personne doivent être respectés à toutes les étapes de développement des systèmes d'IA (Ben-Israel *et al.*, 2020, p. 38). Or, bien que certains droits fondamentaux aient notamment pour objectif de protéger la diversité culturelle, une telle mention demeure trop vague et, en l'absence de précisions, ne produira que peu d'effets favorables à la préservation de la diversité culturelle.

Enfin, soulignons que ces cadres sont des déclarations⁴³ d'intention, non contraignantes, qui guident uniquement les organisations qui les ont

39. *Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*, Université de Montréal, 2018.

40. *Ibid.*, au principe 7 (nos italiques).

41. *Ibid.*, au paragr. 18.

42. *Ibid.*, au paragr. 6.

43. Celles-ci doivent être distinguées des déclarations entendues en tant qu'instruments juridiques adoptés par des États, telles que la *Déclaration universelle sur la diversité culturelle*

adoptées (Donahoe, 2019, p. 118). Plus encore, il s'agit essentiellement de cadres éthiques émanant d'acteurs issus des pays développés (Ben-Israel *et al.*, 2020, p. 37) et ils sont donc peu représentatifs des préoccupations et des valeurs des pays du Sud et des groupes minoritaires (Langlois et Régis, 2021, p. 152). Ces cadres éthiques sont donc « culturellement spécifiques » (Jobin *et al.*, 2019 dans Langlois et Régis, 2021, p. 163), ce qui peut être préjudiciable pour la diversité culturelle dans la mesure où ils tendent à imposer des conceptions et des valeurs représentatives de certaines cultures à d'autres cultures.

Ainsi, le principe de protection de la diversité culturelle est généralement peu présent dans les cadres éthiques. Par ailleurs, selon les instruments, cette notion semble renvoyer à des réalités différentes, ce qui contribue à créer de la confusion, et nécessiterait la clarification du concept de diversité culturelle utilisé. De même, malgré quelques références pertinentes, la portée de tels instruments paraît fortement limitée notamment, car ils n'émanent pas des États eux-mêmes. Il serait alors opportun d'envisager l'éventuelle adoption d'une réglementation en matière d'IA (Conseil de l'Europe, 2020, p. 22 ; Langlois et Régis, 2021, p. 157) impliquant une large diversité d'acteurs, et plus particulièrement des États, afin de protéger plus adéquatement la diversité culturelle.

2.1.2 La Recommandation de l'UNESCO sur l'éthique de l'IA : un instrument juridique opportun pour la préservation de la diversité culturelle

En 2019, l'UNESCO a entrepris l'élaboration d'un instrument normatif international sur l'éthique de l'IA prenant la forme d'une recommandation⁴⁴. Reconnue pour ses travaux en matière d'éthique, l'Organisation offre sans doute le cadre le plus abouti actuellement.

Adoptée le 23 novembre 2021, à l'occasion de la 41^e session de la Conférence générale de l'UNESCO, la *Recommandation sur l'éthique de l'intelligence artificielle* accorde une place fondamentale à la culture en tant que « domaine d'action stratégique »⁴⁵ (Guèvremont *et al.*, 2020). Dès ses considérants, la Recommandation reconnaît la menace que peuvent constituer les technologies de l'IA sur la diversité culturelle et mentionne

(2001) de l'UNESCO (voir *supra* 1.2.2.).

44. À la différence d'une convention, une recommandation n'est pas sujette à ratification, mais les États sont invités à l'appliquer.

45. *Recommandation sur l'éthique de l'intelligence artificielle*, Doc off UNESCO, Doc UNESCO SHS/BIO/PI/2021/1 (2021).

la Convention de 2005 et plusieurs instruments pertinents. Parmi ses objectifs, on peut souligner celui de «protéger, promouvoir et respecter les droits de l’homme et les libertés fondamentales, la dignité humaine et l’équité, [...] ; protéger les intérêts des générations présentes et futures ; [...] et respecter la diversité culturelle à tous les stades du cycle de vie des systèmes d’IA»⁴⁶. La Recommandation édicte plusieurs principes, dont celui d’équité et de non-discrimination et précise que «[l]es États membres devraient s’efforcer de promouvoir un accès inclusif pour tous, y compris pour les communautés locales, à des systèmes d’IA offrant des contenus et des services adaptés localement, et respectant le multilinguisme et la diversité culturelle»⁴⁷. Ces systèmes devraient également «favoriser une offre diversifiée et une pluralité d’accès en matière d’expressions culturelles, en faisant notamment en sorte que la recommandation algorithmique améliore la visibilité et la découvrabilité des contenus locaux» et ce, en se tournant vers les entreprises spécialisées dans les technologies⁴⁸. On constate donc que la diversité culturelle est une préoccupation fondamentale. La place occupée par ces questions n’a néanmoins rien de surprenant étant donné le mandat de l’UNESCO (consacré à l’éducation, la science et la culture) et elle aurait probablement été différente si une autre organisation internationale avait été chargée d’un tel processus.

Par ailleurs, le choix d’un instrument de droit souple (*soft law*) est intéressant dans la mesure où il prévoit une série d’engagements pris par les États au regard de l’IA, tout en reprenant les principes et les fondements juridiques établis par le droit international. Comme le rappellent Langlois et Régis, la forme que devrait prendre un éventuel instrument sur l’IA est un débat en soi, mais il n’en demeure pas moins que la forme de ces futures normes – qu’elles soient contraignantes ou non – importe peu puisqu’elles peuvent toutes deux avoir un effet «*even if they do not address all the potential opportunities and problems that this thriving innovation will raise*» (Langlois et Régis, 2021, p. 164). En effet, les normes contraignantes (*hard law*) et non contraignantes (*soft law*) influencent le droit national (Langlois et Régis, 2021, p. 164 ; Varin et Guèvremont, 2020). La mise en place d’un instrument non contraignant semble néanmoins plus adaptée actuellement dans la mesure où l’élaboration d’instruments contraignants demeure «*long, complex, rigid and costly*» et donc «*less suitable for issues that*

46. Recommandation sur l’éthique de l’intelligence artificielle, *ibid.*, au paragr. 8 (c).

47. *Ibid.*, au paragr. 28.

48. *Ibid.*, au paragr. 98.

require a rapid response and flexibility to adapt to scientific developments [...] » (Langlois et Régis, 2021, p. 165).

À la différence des instruments élaborés jusqu'à présent, pour la plupart par une multitude de parties prenantes (étatiques ou non, à l'échelle régionale, etc.), cette Recommandation rassemble la quasi-totalité des États. Un tel aspect est fondamental en raison des atteintes possibles des systèmes d'IA sur l'ensemble des droits protégés en droit international. En effet, un encadrement de cette technologie par les acteurs privés, ou même la société civile, ne peut s'avérer effectif sans la participation des États qui sont responsables de la réalisation de ces droits ; l'inverse est également vrai. Cependant, bien que la Recommandation s'adresse aux « États membres, à la fois en tant qu'acteurs de l'IA et en tant qu'autorités responsables de l'élaboration de cadres juridiques et réglementaires tout au long du cycle de vie des systèmes d'IA, et de la promotion de la responsabilité des entreprises », elle s'adresse également à tous les acteurs de l'IA issus du secteur public ou privé en leur fournissant des orientations éthiques⁴⁹.

Enfin, il convient de noter que cette Recommandation unit l'éthique et le droit. Une approche combinée que suggèrent d'ailleurs plusieurs auteurs pour qui aborder l'IA exclusivement sous l'angle de l'éthique ou du droit n'est pas suffisant (Fierens *et al.*, 2021). En effet, bien que l'éthique puisse guider vers une « IA morale », les auteurs considèrent que des règles contraignantes sont essentielles « *for incentivising AI stakeholders to engage in AI solutions that maximise the benefits and minimise the harms of AI to individuals and society* ». L'éthique influence ainsi le droit dans une direction ou une autre (Fierens *et al.*, 2021, p. 50 et 71).

2.2 Vers une IA culturellement responsable ?

Les normes internationales permettent, entre autres, de répondre à des problématiques qui ne peuvent être résolues par un seul pays ou qui dépassent les frontières étatiques (Langlois et Régis, 2021, p. 162). Pour le Groupe de haut niveau sur la coopération numérique des Nations Unies, des « *collaborative international governance tools* » sont nécessaires pour l'IA en raison de ses effets sur l'humanité (Langlois et Régis, 2021, p. 162).

Ces normes doivent reposer sur les instruments internationaux déjà existants qui façonnent le droit international. Jusqu'à présent, on constate

49. Recommandation sur l'éthique de l'intelligence artificielle, *supra* note 45, au paragr. 4.

que les droits de la personne semblent être le prisme d'entrée pour régler ces systèmes (Ben-Israel *et al.*, 2020, p. 81). Or, cet angle peut s'avérer imprécis en ce qui concerne la diversité culturelle, ce qui nécessite des références plus précises. Le Conseil de l'Europe offre une piste intéressante en affirmant que « les valeurs telles que la protection des minorités, le pluralisme et la diversité devraient absolument être examinées dans le cadre du développement de ces solutions » (Ben-Israel *et al.*, 2020, p. 81). Par ailleurs, les objectifs et les principes de certains instruments du droit international, notamment de la Convention de 2005, paraissent primordiaux. Les travaux menés ces dernières années permettant d'ores et déjà d'identifier les atteintes portées à la diversité des expressions culturelles par ces systèmes d'IA, et donc les éléments à prendre en considération.

Cette contribution ne vise pas à résoudre les défis de l'encadrement de l'IA, mais à démontrer l'importance pour les initiatives prises en ce sens d'être plus inclusives et de tenir compte de la diversité culturelle. Pour ce faire, la promotion d'une IA *culturellement* responsable, c'est-à-dire qui respecte la diversité des cultures, semble être une voie à explorer. Cette diversité ne doit pas se limiter à la représentation d'une diversité dans les équipes de conception des systèmes, comme cela est souvent le cas, mais constituer une fin en soi qui doit apparaître dans les résultats desdits systèmes. L'IA aurait donc le potentiel d'exposer les individus à davantage de diversité, si cet aspect est reconnu et intégré dans la conception de ces systèmes (*by design*). Ainsi, en prônant une IA *culturellement* responsable, les États seraient davantage incités à respecter leurs engagements envers les instruments internationaux dont ils sont parties.

Promouvoir une IA *culturellement* responsable pourrait prendre plusieurs formes, mais il convient tout d'abord de définir ce qui est entendu par « responsable », un terme largement utilisé par les instruments visant à proposer un cadre éthique pour l'IA. La définition proposée par la Recommandation de l'UNESCO semble appropriée dans la mesure où elle fait référence aux droits de la personne. Elle précise en effet que :

Les acteurs de l'IA et les États membres devraient respecter, protéger et promouvoir les droits de l'homme et les libertés fondamentales, [...] endossant leur responsabilité éthique et juridique respective, conformément au droit national et international, en particulier les obligations des États membres en matière de droits de l'homme, et aux directives éthiques établies tout au long du cycle de vie des systèmes d'IA, y compris en ce qui concerne les acteurs de l'IA sur leur territoire et sous leur contrôle effectifs. La responsabilité éthique des décisions et actions fondées d'une quelconque manière

sur un système d'IA devrait toujours incomber en dernier ressort aux acteurs de l'IA selon le rôle qui est le leur dans le cycle de vie du système d'IA⁵⁰.

En lien avec cette définition, quelques pistes de solutions pourraient être envisagées afin de favoriser le développement d'une IA *culturellement* responsable. La première serait la prise en considération, lors de l'élaboration d'instruments relatifs à l'IA par les États, des orientations déjà adoptées afin de favoriser le respect de leurs engagements internationaux dans l'environnement numérique (voir la section 1.2). Par exemple, l'ajout de dispositions inspirées de celles qui sont contenues dans les *Directives opérationnelles pour la mise en œuvre de la Convention de 2005 dans l'environnement numérique* visant à garantir la visibilité et la découvrabilité des contenus culturels nationaux et locaux⁵¹ ou encore à assurer l'accès libre et pérenne à diverses expressions culturelles⁵², pourrait être pertinent (Kulesz, 2018, p. 13). Par ailleurs, l'adoption de réglementations sectorielles de l'IA propres à la culture pourrait être envisagée à plusieurs échelles (national, international, etc.). Complémentaires à une réglementation générale, elles permettraient de clarifier les responsabilités des États et des acteurs privés en matière de diversité culturelle (Conseil de l'Europe, 2020, p. 27). En outre, soulignons qu'une obligation d'évaluation des systèmes d'IA devrait être appliquée systématiquement afin de garantir qu'ils respectent les droits de la personne (Ben-Israel *et al.*, 2020, p. 8) et favorisent donc le respect de la diversité culturelle. Un tel mécanisme, idéalement sous l'égide d'une autorité indépendante nationale ou régionale, devrait évaluer les impacts négatifs de tels systèmes et proposer des mesures pour les éviter ou les atténuer dès lors qu'ils seront déployés sur le territoire des États (Pineschi, 2012, p. 45). Enfin, une dernière piste de solutions pouvant être envisagée serait de sensibiliser aux enjeux culturels, les négociateurs qui participent à l'adoption d'instruments relatifs à l'IA, mais également les comités d'experts indépendants consultés. En effet, les spécialistes des droits humains généralement mobilisés sont peu sensibles aux enjeux liés à la protection des droits culturels et de la diversité culturelle. De même, les spécialistes du droit de la culture ne disposent pas toujours de l'expertise requise pour comprendre l'impact des systèmes d'IA sur la diversité culturelle.

50. *Recommandation sur l'éthique de l'intelligence artificielle*, supra note 45 au para 42.

51. *Directives opérationnelles sur la mise en œuvre de la Convention dans l'environnement numérique*, supra note 26 au paragr. 16.1.

52. *Ibid.*, au paragr. 17.

En conclusion, la présente contribution vise à explorer des pistes de solutions afin d'assurer, dans la mesure du possible, le développement de systèmes d'IA *culturellement* responsables. Dans un monde de plus en plus numérisé, soumis aux diktats des algorithmes et autres systèmes dont l'IA, les États ont la responsabilité de s'assurer que leurs engagements pris en matière de diversité culturelle perdurent pour les générations futures. Les bienfaits de la protection de la diversité culturelle ont été prouvés et face à l'IA cet enjeu demeure plus que jamais d'actualité pour l'avenir de nos sociétés. Si la diversité culturelle apparaît parfois dans les cadres développés aujourd'hui en la matière, il est impératif qu'elle soit désormais incluse dans l'ensemble des instruments qui visent à réglementer l'IA, à l'échelle à la fois internationale, régionale et nationale. La *Recommandation de l'UNESCO sur l'éthique de l'IA* est un premier pas favorable en ce sens, mais il ne peut demeurer le seul.

RÉFÉRENCES

Instruments juridiques

Traités

Pacte international sur les droits économiques, sociaux et culturels, 999 RTNU 3, 16 décembre 1966 (entrée en vigueur : 3 janvier 1976) [PIDESC].

Pacte international relatif aux droits civils et politiques, 999 RTNU 171, 19 décembre 1966 (entrée en vigueur : 23 mars 1976) [PIDCP].

Convention sur la protection et la promotion de la diversité des expressions culturelles, 2440 RTNU 311, 20 octobre 2005 (entrée en vigueur : 18 mars 2007) [Convention de 2005].

Déclarations

Déclaration universelle des droits de l'homme, Rés AG 217A (III), Doc off AGNU, 3^e sess, supp no. 13, Doc NU A/810 (1948) [DUDH].

Déclaration de Mexico sur les politiques culturelles, adoptée lors de la Conférence mondiale sur les politiques culturelles du 26 juillet au 6 août 1982, UNESCO, Mexico, Mexique (1982) [Déclaration de Mexico].

Déclaration universelle de l'UNESCO sur la diversité culturelle, Rés CG, Doc off UNESCO, 31^e sess, Doc UNESCO CLT-2002/WS/9 (2001).

Recommandations

UNESCO, *Recommandation sur l'éthique de l'intelligence artificielle*, SHS/BIO/PI/2021/1, 23 novembre 2021.

Droit dérivé

Observation générale n° 21 : Droit de chacun de participer à la vie culturelle (Article 15, paragraphe 1.a), Doc off Comité des droits économiques, sociaux et culturels, 43^e sess, Doc NU E/C.12/GC/21 (2009).

Directives opérationnelles sur la mise en œuvre de la Convention dans l'environnement numérique, approuvées par la Conférence des Parties à la Convention de 2005 sur la protection et la promotion de la diversité des expressions culturelles lors de sa sixième session, Paris (2017).

Instruments non juridiques

Asilomar AI Principles, Future of Life Institute, 2017.

Déclaration de Toronto : Protéger le droit à l'égalité et à la non-discrimination dans les systèmes reposant sur l'apprentissage automatique, Amnesty International et Access Now, Canada, 2018.

Déclaration de Fribourg sur les droits culturels, Groupe de Fribourg, 2007 [Déclaration de Fribourg].

Lignes directrices en matière d'éthique pour une IA digne de confiance, Commission européenne, Groupe d'experts de haut niveau sur l'intelligence artificielle (GEHN IA), 8 avril 2019.

Déclaration de Montréal pour un développement responsable de l'intelligence artificielle, Université de Montréal, 2018.

Doctrine

Ben-Israel, Issac et al., « Vers une régulation des systèmes d'IA : Perspectives internationales sur l'élaboration d'un cadre juridique fondé sur les normes du Conseil de l'Europe dans le domaine des droits de l'homme, de la démocratie et de l'État de droit », Compilation de contributions DGI (2020)16 préparé par le Secrétariat du CAHAI, Strasbourg, Conseil de l'Europe, DGI (2020)16, décembre 2020.

Bennoune Karima, *Rapport de la Rapporteuse spéciale dans le domaine des droits culturels « L'importance des espaces publics pour l'exercice des droits culturels »*, Doc off Conseil des droits de l'homme, 74^e sess, Doc NU A/74/255, 2019.

Bensamoun, Alexandra et Joëlle Farchy, *Mission Intelligence artificielle et culture*, Rapport final, Conseil supérieur de la propriété littéraire et artistique, 27 janvier 2020.

Bernier, Ivan et Hélène Ruiz Fabri, *Évaluation de la faisabilité juridique d'un instrument international sur la diversité culturelle*, Groupe franco-québécois sur la diversité culturelle, Gouvernement du Québec, Ministère de la Culture, 2002.

Bernier, Ivan et Dave Atkinson, « Mondialisation de l'économie et diversité culturelle : les arguments en faveur de la préservation de la diversité culturelle », Document de réflexion, 2000.

Blake, Janet, *Exploring Cultural Rights and Cultural Diversity: An Introduction with Selected Legal Materials*, Builth Wells, Institute of Art and Law, 2014.

Burri, Mira, « Découvrabilité de contenus locaux, régionaux et nationaux en ligne », Document de réflexion pour le ministère du Patrimoine canadien et la Commission canadienne pour l'UNESCO, 2019.

Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST), *Étude préliminaire sur l'Éthique de l'intelligence artificielle de la COMEST*, SHS/COMEST/EXTWG-ETHICS-AI/2019/1, Paris, UNESCO, 26 février 2019.

- Conseil de l'Europe, « Un cadre juridique pour les systèmes d'intelligence artificielle – Étude de faisabilité d'un cadre juridique pour le développement, la conception et l'application de l'intelligence artificielle, basé sur les normes du Conseil de l'Europe sur les droits de l'homme, la démocratie et l'État de droit », adopté par le CAHAI lors de sa 3^e réunion plénière, Strasbourg, Conseil de l'Europe, DGI (2021)04, 17 décembre 2020.
- Daelman, Charline, « AI through a human rights lens. The role of human rights in fulfilling AI's potential », dans Jan De Bruyne et Cedric Vanleenhove (dir.), *Artificial Intelligence and the Law*, Cambridge, Intersentia, 2021.
- Donahoe, Eileen et Megan Metzger MacDuffee, « Artificial intelligence and human rights », *Journal of Democracy*, vol. 30, n° 2, 2019, p. 115-126.
- Donders Yvonne, « Cultural diversity and cultural identity » dans Andreas J. Wiesand, Kalliopi Chainoglou, Anna Sledzinska-Simon avec Yvonne Donders (dir.), *Culture and Human Rights: The Wroclaw Commentaries*, Berlin/Boston, De Gruyter, 2016, p. 23-32.
- Fierens, Michiel, Stephanie Rossello et Ellen Wauters, « Setting the scene : On AI ethics and regulation », dans Jan De Bruyne, Cedric Vanleenhove (dir.), *Artificial Intelligence and the Law*, Cambridge, Intersentia, 2021.
- Gentelet, Karine et Alexandra Bahary-Dionne, « Stratégies des Premiers Peuples au Canada concernant les données numériques : décolonisation et souveraineté », *tic&société*, vol. 15, n° 1, 2021, p. 189-208.
- Guèvremont, Véronique (dir.) *et al.*, « Les mesures de la découvrabilité des contenus culturels francophones dans l'environnement numérique : compte rendu des tendances et recommandations », en collaboration avec Octavio Kulesz, Québec, Chaire UNESCO sur la diversité des expressions culturelles, 2019.
- Guèvremont, Véronique, Clémence Varin et Iris Richer, « Commentaires de la Chaire UNESCO sur la diversité des expressions culturelles sur l'Avant-projet de recommandation sur l'éthique de l'intelligence artificielle de l'UNESCO », remis au Ministère de la Culture et des Communications du Québec, 2 décembre 2020.
- Guèvremont, Véronique et Maxime Mariage, « Fair culture – A key to sustainable development », Bonn, German Commission for UNESCO, 2021.
- Hunt, Robert et Fenwick McKelvey, « Algorithmic regulation in media and cultural policy : A framework to evaluate barriers to accountability », *Journal of Information Policy*, vol. 9, 2019, p. 307-335.
- Institut Montaigne, *Algorithmes : contrôle des biais S.V.P.*, Paris, 2020.
- Kulesz, Octavio, « Culture, machines et plateformes : l'intelligence artificielle et son impact sur la diversité des expressions culturelles », Comité intergouvernemental pour la protection et la promotion de la diversité des expressions culturelles, douzième session, Paris, Siège de l'UNESCO, 11-14 décembre 2018, DCE/18/12.IGC/INF.4, 13 novembre 2018.
- Kulesz, Octavio et Thierry Dutoit, *L'intelligence artificielle dans l'art et les industries culturelles et créatives. Panorama des technologies, expertises et bonnes pratiques dans l'espace francophone*, Paris, OIF, 2020.
- Langlois, Lyse et Catherine Régis, « Analyzing the contribution of ethical charters to building the future of artificial intelligence governance », dans Bertrand Braunschweig et Malik Ghallab (dir.), *Reflections on Artificial Intelligence for Humanity*, Cham, Springer, 2021.
- Lewis, Jason Edward *et al.*, *Indigenous Protocol and Artificial Intelligence Position Paper*, Indigenous Protocol and Artificial Intelligence Working Group and the Canadian Institute for Advanced Research, Honolulu, 2020.

- McKelvey, Fenwick et Robert Hunt, « Responsabilité algorithmique et découverte de contenus numériques », Document de réflexion pour le ministère du Patrimoine canadien et la Commission canadienne pour l'UNESCO, 2019.
- Pineschi, Laura, « Cultural diversity as a human right? General comment no. 21 of the Committee on economic, social and cultural rights », dans Silvia Borelli, Federico Lenzerini, *Cultural Heritage, Cultural Rights, Cultural Diversity: New development in international law*, Leiden, Martinus Nijhoff Publishers, 2012.
- Raso, Filippo, Hannah Hilligoss, Vivek Krishnamurthy, Christopher Bavitz, Kim Levin, « Artificial intelligence and human rights: Opportunities and risks », *Berkman Klein Center for Internet & Society Research Publication*, 25 septembre 2018.
- Romainville Céline, « Participation in cultural life », dans Andreas J. Wiesand, Kalliopi Chainoglou, Anna Sledzinska-Simon avec Yvonne Donders (dir.), *Culture and Human Rights: The Wroclaw Commentaries*, Berlin/Boston, De Gruyter, 2016.
- Shaheed, Farida, *Rapport de l'Experte indépendante dans le domaine des droits culturels soumis en application de la résolution 10/23 du Conseil des droits de l'homme*, Doc off Conseil des droits de l'homme, 14e sess, Doc NU A/HRC/14/36 (2010).
- Varin, Clémence et Véronique Guèvremont, « La Convention sur la diversité des expressions culturelles quinze ans après : une mise en œuvre effective d'un instrument juridique faiblement contraignant », *Revue québécoise de droit international*, n° 32.2-2019, 10 décembre 2020.

ALGORITHMES ET PROPAGATION DE NORMES CULTURELLES SEXOSPÉCIFIQUES

Eleonore Fournier-Tombs¹ et Céline Castets-Renard²

L'intelligence artificielle est utilisée de plus en plus pour créer des interfaces technologiques – que ce soit des agents conversationnels (*chatbots*), des assistants personnels ou des robots ayant comme fonction d'interagir avec des êtres humains. Ces interfaces offrent des services, répondent à des questions et entreprennent même des tâches domestiques, telles que l'achat de produits d'épicerie ou le contrôle de la température de la maison.

Dans une étude sur les assistants personnels ayant des voix féminines, tels que Alexa d'Amazon et Siri d'Apple, l'Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO) a fait valoir que ces technologies pouvaient avoir des effets négatifs considérables sur la promotion de l'égalité des sexes (UNESCO *et al.*, 2019). Outre le fait que ces systèmes d'intelligence artificielle (IA) soient formés sur des modèles sexospécifiques, ces assistants à voix féminines présentent tous des modèles féminins stéréotypés. Ce problème est aggravé par le fait que ces systèmes ont probablement été créés principalement par des développeurs masculins (Simonite, 2018). Ces assistants sexospécifiques peuvent constituer une menace par la représentation biaisée des femmes qu'ils génèrent, d'autant qu'ils sont de plus en plus omniprésents dans nos vies quotidiennes.

1. Directrice de la recherche sur l'action anticipatoire et l'innovation au Centre de recherche politique des Nations Unies à New York.

2. Professeure titulaire à la Faculté de droit civil à l'Université d'Ottawa.

On prévoit ainsi qu'à partir de 2021 il y aura davantage d'assistants vocaux que d'êtres humains sur la planète (UNESCO et EQUALS Skills Coalition, 2019, p. 3).

Compte tenu de l'utilisation croissante d'assistants vocaux à voix féminine et formés à l'aide de modèles linguistiques biaisés, l'effet potentiel sur les normes de genre est préoccupant. Alors que l'isolement a considérablement augmenté pendant la COVID-19, il existe un risque que la principale interaction « féminine » de certaines personnes se fasse avec ces assistants vocaux. Si l'on n'y prend garde, des représentations sexistes des femmes, en décalage total avec des femmes réelles, vont s'immiscer au cœur de l'intimité du domicile ou de nos téléphones intelligents, en tous lieux et tous temps³. Au demeurant, les modèles sont essentiellement les mêmes, conduisant à reproduire un « standard » unique et un lissage culturel dans l'interaction homme-machine, niant la diversité des utilisateurs de ces produits dans le monde.

Si d'aucuns ont fait valoir que les algorithmes d'apprentissage peuvent être moins biaisés que les êtres humains, souvent influencés par des normes culturelles discriminatoires dont ils n'ont pas nécessairement conscience (Miller, 2018), ce constat est sans égard pour le fait que l'intelligence artificielle (IA) est nécessairement créée par des êtres humains dont elle intègre la façon de raisonner. Il est en effet facile de sous-estimer l'importance des normes culturelles dans la prise de décision humaine. L'intelligence artificielle imite les préjugés sociaux des données qui lui ont été fournies, à moins qu'elle ne soit conçue explicitement avec des principes différents. Il n'est donc pas surprenant qu'une intelligence artificielle développée sans valeurs intégrées ne fasse que refléter des normes sociales déjà biaisées.

Ce chapitre étudie l'effet ambigu des algorithmes d'apprentissage qui menacent de propager des normes de genre biaisées, tout en promettant de les éliminer. Deux principales questions sont ici posées : comment mieux connaître et documenter les biais de genre qui peuvent être associés à d'autres formes de domination ? Comment mieux utiliser la technique pour en atténuer les effets négatifs ?

Les objectifs de la recherche sont doubles. Il s'agit, d'une part, d'analyser les préjugés culturels ancrés dans les modèles d'apprentissage automatique,

3. 88 % des Canadiens de plus de 15 ans qui utilisent Internet possèdent un téléphone intelligent, et 56 % de ce groupe dorment avec leur téléphone à portée de main, selon Statistique Canada : <https://www150.statcan.gc.ca/t1/tbl1/fr/tv.action?pid=2210011501&request_locale=fr>.

tels que les associations de mots sexistes et les représentations féminines stéréotypées, suivant différentes perspectives culturelles et géographiques (1). D'autre part, cette étude propose de considérer les moyens de l'apprentissage automatique, afin d'inverser les normes biaisées liées au genre, par la fourniture de prédictions et décisions exemptes de discrimination systémique ou par la minimisation des discriminants négatifs (2).

1. PROPAGATION DE NORMES DISCRIMINATOIRES ENVERS LES FEMMES DANS DES ALGORITHMES BIAISÉS

1.1 Représentation des femmes dans l'IA

De façon générale, les techniques de traitement du langage naturel peuvent être particulièrement révélatrices des biais de genre. Les grands modèles linguistiques qui constituent une catégorie de ces techniques permettent de mettre en correspondance les mots avec des vecteurs numériques, en tenant compte de leur contexte dans les textes volumineux. Ces techniques sont souvent utilisées pour prédire le mot suivant dans une séquence, ce qui permet aux développeurs de créer des agents conversationnels réalistes ou encore des articles de blogues générés par ordinateur, etc. La plupart des modèles disponibles sont fournis avec une formation préalable, ce qui signifie que les mots et leurs associations ont généralement été cartographiés à partir de textes disponibles en ligne, comme Wikipédia et Google Books (Devlin *et al.*, 2018).

Dans un document saisissant datant de 2016, Bolukbasi et ses collaborateurs (2016) ont constaté que les associations intégrées dans ces modèles contenaient de forts stéréotypes de genre. Par exemple, les prénoms féminins étaient associés à des termes tels que femme au foyer, infirmière et réceptionniste, tandis que les prénoms masculins étaient associés à des termes tels que philosophe, capitaine et architecte. Concrètement, cela signifie que, dans les technologies qui utilisent de tels modèles, le texte ou la conversation générée par l'ordinateur fait des suppositions très sexistes.

Plus encore, lorsqu'il s'agit de qualifier les femmes noires dans des suggestions des moteurs de recherche comme Google Search, Noble a montré que les préjugés sont nombreux et particulièrement dénigrants (Noble, 2018). Les biais des données racistes s'ajoutent alors aux biais de genre (Benjamin, 2019). Buolamwini montre aussi dans sa thèse réalisée au MIT Media Lab que le taux d'erreur des outils de reconnaissance faciale est plus élevé envers les femmes que les hommes, envers les peaux sombres

que les peaux claires. Les femmes aux peaux sombres sont alors moins bien reconnues par le système (Buolamwini *et al.*, 2018). Pour autant, les travaux sur l'intersectionnalité montrent la complexité et la spécificité des discriminations envers les femmes issues des minorités ou communautés racisées qui ne peuvent être décrites par une simple addition des problèmes de chaque groupe dominé (Crenshaw, 1989). Dans ce contexte, les données utilisées dans certains systèmes d'IA ne font qu'exacerber les problèmes bien connus que les femmes doivent affronter depuis longtemps. Cependant, ces problèmes sont alors systématisés et renforcés.

Par ailleurs, les systèmes d'IA peuvent aussi ne pas être correctement renseignés. L'intelligence artificielle se nourrit de données et les questions fondamentales à l'endroit des femmes vont être de savoir de quelle qualité mais aussi quantité de données dispose-t-on pour révéler les problèmes liés au genre. Les travaux menés par exemple par Maria Salguero pour révéler les informations sur les féminicides au Mexique et prouver l'ampleur du phénomène illustrent parfaitement les enjeux de dissimulation qui se perpétuent dans les bases de données⁴. Certaines informations sont volontairement cachées ou ignorées, ce qui pose la question de savoir pour qui sont faites les sciences des données et dans quel but ou pour défendre quel intérêt. L'absence de données a pour effet de rendre invisibles les agressions faites aux femmes et, plus largement, aux minorités opprimées ou racisées.

Il est alors à craindre que la systématisation et l'automatisation des clichés, biais ou dissimulations, sous couvert d'une apparente neutralité mathématique, rendent encore plus difficile voire impossible la possibilité de les connaître et donc de les corriger. Au contraire, leur révélation doit permettre de rechercher des solutions dans les systèmes algorithmiques eux-mêmes.

1.2 La féminisation de l'IA

Comprendre les préjugés sexistes dans l'IA implique aussi de décortiquer le concept de « féminisation de l'IA ». Ces dernières années, un certain nombre d'objets d'IA couramment utilisés se sont ainsi vu attribuer un « genre féminin ». Il s'agit notamment des agents conversationnels, des

4. *Los Feminicidios en Mexico*, « Los Feminicidios en Mexico Data Platform », 2021, <<https://feminicidiosmx.crowdmap.com/main>> (consulté le 1^{er} avril 2021).

robots (comme le robot Sophia) et des assistants vocaux (comme ALICE⁵). Ce genre féminin est attribué de plusieurs manières différentes, à savoir l'apparence, la voix, la personnalité et le langage. Le tableau ci-dessous montre comment ces quatre attributs sont utilisés pour construire un objet d'IA de sexe féminin.

TABLEAU 20.1

Attributs féminisés des objets d'IA

OBJET	APPARENCE	VOIX	PERSONNALITÉ	LANGAGE
Assistant vocal	Pas d'attributs féminins	Voix féminine par défaut	Complaisant, séducteur	Entraîné sur un modèle de langage stéréotypé ⁶
Robot (Sophia)	Jeune femme blanche, visage, corps aux proportions féminines	Voix féminine	Complaisant, séducteur	Entraîné sur un modèle de langage stéréotypé
Agent conversationnel	Variable (selon les implantations)	Pas de voix genrée	Souvent neutre	Entraîné sur un modèle de langage stéréotypé

Il est communément admis que ces objets d'IA ont été développés avec des attributs féminins parce que leur principale fonction est l'assistance humaine. Les femmes ont historiquement été beaucoup plus susceptibles d'être considérées comme des assistantes, avec des traits de personnalité tels que la serviabilité, la conformité, la subordination et une petite mesure de séduction encouragée (Criado Perez, 2019). Même récemment, une étude du recensement américain de 2010 a révélé que les emplois les plus courants pour les femmes n'ont pas changé depuis 1950. En 1950, ces emplois étaient les suivants : secrétaire, caissière de banque, vendeuse, employée de maison et enseignante. En 2010, il s'agissait des emplois suivants : secrétaire, caissière, enseignante dans les écoles primaires et secondaires, infirmière et aide-soignante (Kurtz, 2013).

-
5. ALICE, qui est l'acronyme d'*Artificial Linguistic Internet Computer Entity*, a été conçu et lancé par Dr Richard Wallace en 1995. ALICE aurait servi d'inspiration pour le système d'exploitation compagnon dans le film de science-fiction *Her*, réalisé par Spike Jonze en 2013.
 6. Cela sera détaillé davantage plus loin dans le texte.

Des entreprises comme Apple et Google ont affirmé que l'attribution de traits féminins aux objets d'IA était entièrement dictée par la demande. En effet, dans les groupes de discussion, les utilisateurs ont simplement mieux réagi aux objets ayant des attributs féminins qu'à ceux ayant des attributs masculins (UNESCO et EQUALS Skills Coalition, 2019). Mais une telle affirmation est par nature elle-même biaisée par les caractéristiques du groupe social le plus susceptible de consommer ces technologies. Les jeunes hommes sont particulièrement friands de ces technologies et constituent les principaux clients à qui s'adressent les entreprises qui les développent (King, 2015).

1.3 L'omniprésence de l'IA féminisée

Pourtant, les assistants vocaux en particulier sont omniprésents. Un rapport de 2017 du Pew Research Center révèle par exemple que 46 % des adultes américains utilisent un assistant vocal numérique, principalement sur leurs téléphones intelligents (Pew Research Center, 2017). En 2020, un rapport de NPR a constaté que la pandémie de COVID-19 avait considérablement augmenté l'utilisation des assistants vocaux, passant de 46 % à 52 % des adultes américains au cours de la crise sanitaire (National Public Media, 2020).

Le rapport de 2019 publié par Microsoft sur les assistants vocaux présente des chiffres encore plus élevés, selon lesquels 72 % des personnes interrogées utiliseraient la recherche vocale au moyen d'un assistant numérique personnel, 35 % la recherche vocale par une enceinte domestique intelligente et 52 % des commandes vocales par une enceinte domestique (vraisemblablement autres que la recherche) (Microsoft, 2019). Les assistants vocaux inclus dans l'enquête de Microsoft – Cortana, Siri, Google Assistant et Alexa – ont tous une voix féminine et communiquent à l'aide de modèles linguistiques formés par l'IA.

Des chercheurs ont également constaté que les êtres humains sont fortement influencés par leur environnement, lorsqu'il s'agit de modifier leurs perceptions. En d'autres termes, si les gens sont constamment confrontés à des représentations stéréotypées des femmes, ils finissent par les croire et les intérioriser. À titre d'exemple, Steele a étudié l'influence de l'exposition à des images stéréotypées féminines (telles que des fleurs, des bijoux et des poupées) sur la perception des femmes de leurs propres

aptitudes en mathématiques, et a découvert que plus elles voyaient ces images, moins elles se sentaient compétentes (Steele, 2003).

1.4 Le langage genré : toujours et encore...

Bien que l'étude de Bolukbasi et ses collaborateurs sur les modèles linguistiques genrés date de 2016, les compagnies les développant ont été très lentes à adopter le changement. Aujourd'hui, il existe, par exemple, de nombreuses applications qui offrent des services de traduction automatisée avec l'apprentissage machine. Or, la fréquence des biais de genre y est toujours frappante. On peut observer cet écart particulièrement lors de la traduction de phrases dans une langue n'utilisant pas de pronoms genrés, tel que le pinyin (qui est le mandarin utilisant des caractères romains, plutôt que des caractères chinois). Ci-dessous, on peut voir la traduction automatisée de deux logiciels très courants, Google Translate et Microsoft Bing, effectuée en mars 2021, soit cinq ans après les découvertes de Bolukbasi et ses collaborateurs.

L'étude a été menée en introduisant le pinyin dans les deux traducteurs. Les résultats sont ensuite présentés traduits en français et en anglais. Dans certains cas cependant, l'inexactitude de la traduction peut être attribuée au fait que le pinyin est tapé sans accents, alors qu'en mandarin les voyelles peuvent avoir quatre accents différents, chacun ayant une signification différente⁷.

Nous pouvons voir ici que Google Translate et Microsoft Bing ont été formés sur deux modèles de langage différents. Le modèle linguistique de Google suppose le genre en fonction du nom ou du verbe : investir est pour l'homme ce que danser est pour la femme ; programmeur est pour l'homme ce que secrétaire est pour la femme. Ces résultats vont exactement dans le même sens que l'article de Bolukbasi et ses collaborateurs (2016). L'outil Microsoft Bing, pour sa part, attribue par défaut le genre masculin dans tous les cas. Le genre féminin est soit utilisé dans un contexte subordonné, ou dans la sphère privée, soit totalement invisible. En tant que tels, les deux outils sont donc discriminatoires, mais de manière différente.

7. Par exemple, les quatre significations du yuan, qui se perdent lors de la saisie dans un logiciel de traduction, sont **yuán** (origine, ou racine), **yuàn** (vouloir, être prêt), **yuǎn** (il y a longtemps) et **yuān** (unité monétaire chinoise de base). Dans ce contexte, l'utilisateur avait l'intention de taper yuán, mais le logiciel a interprété yuǎn.

TABLEAU 20.2
Choix de pronoms féminins et masculins par les algorithmes courants de traduction

ORIGINAL PINYIN	GOOGLE TRANSLATE – ANGLAIS	GOOGLE TRANSLATE – FRANÇAIS	MICROSOFT BING TRANSLATOR – ANGLAIS	MICROSOFT BING TRANSLATE – FRANÇAIS
ta zhengzai touzi	He is investing	Il investit	He's investing	Il investit
ta shi zhexue jia	He is a philosopher	C'est un philosophe	He is a philosopher	C'est un philosophe
ta shi chengxu yuan	He is a programmer	Il est programmeur	He is a program dollar (see footnote)	C'est un dollar de programme (see footnote)
ta shi mishu	She is a secretary	Elle est secrétaire	He is the secretary	Il est le secrétaire
ta shi duizhang	He is a captain	Il est le capitaine	Unable to translate	Unable to translate
ta zai xi yifu	She is doing laundry	Elle fait la lessive	He's washing clothes	Il lave les vêtements
ta zai zhaogu haizi	She is taking care of the child	Elle prend soin de l'enfant	He's taking care of the kids	Il s'occupe des enfants
ta zai gongzuo	He is working	Il travaille	He's at work	Il est au travail
ta zai kaiche	He is driving	Il conduit	He's driving	Il conduit
ta zai tiaowu	She is dancing	Elle danse	He's dancing	Il danse

L'exercice de traduction est un moyen d'illustrer les biais des modèles linguistiques utilisés par les entreprises technologiques qui développent non seulement les logiciels de traduction, mais aussi les assistants vocaux.

1.5 Le comportement genré des utilisateurs : sexualisation et harcèlement des assistants personnels

Après avoir illustré la manière dont les outils de traduction ou d'assistance vocale répondent aux entrées des utilisateurs, nous analysons ici, réciproquement, la manière dont les utilisateurs interagissent avec ces outils. Sur son site Web d'assistance (Apple, 2021), Apple suggère aux utilisateurs de poser des questions, telles que : « Dis Siri, quel temps fait-il aujourd'hui ? » Cependant, une exploration des termes de recherche et des vidéos YouTube suggère que les utilisateurs s'engagent dans des interactions sexuées beaucoup plus stéréotypées avec le logiciel, au-delà de la météo.

L'étude suivante analyse 20 vidéos trouvées sur YouTube après avoir tapé le terme de recherche suivant : « Top 10 des questions que vous ne devriez jamais poser à Alexa (Siri, Cortana) ». Ces vidéos, extrêmement populaires parmi les producteurs de contenus, ont une intention humoristique, montrant les réponses de ces outils au harcèlement, à la sexualisation et à la domination. Dans presque toutes les vidéos, les producteurs de contenu anthropomorphisent les outils, c'est-à-dire qu'ils utilisent des pronoms sexués lorsqu'ils interagissent avec eux. La composante humoristique émane aussi largement de l'inattendu, par exemple lorsque Siri ne réagit pas comme le ferait une femme humaine. L'objectif de l'étude n'est pas d'avoir une approche exhaustive des réactions genrées des assistants vocaux, mais d'avoir quelques indices sur leur proportion et leurs caractéristiques.

Le tableau ci-dessous représente le nombre de fois, dans une vidéo, que le producteur ou la productrice a posé des questions à un assistant personnel de façon dérogatoire. Ainsi, les vidéos analysées ont en moyenne utilisé 1,84 phrase de harcèlement, 1,11 de sexualisation, 2,58 à connotation romantique et 3,68 fois des ordres. Les vidéos ont été produites entre 2017 et 2020, et ont été codées manuellement par l'équipe de recherche.

TABLEAU 20.3

Résumé du nombre de phrases dérogatoires dans les vidéos observées

	HARCÈLEMENT	SEXUALISATION	ROMANTIQUE	ORDRES
Moyenne	1,84	1,11	2,58	3,68
Médiane	0	0	2	3
Minimum	0	0	0	0
Maximum	13	5	12	14

À titre d'exemple, de nombreux producteurs et productrices demandent à l'assistant virtuel de répondre à des questions telles que : « Voulez-vous sortir avec moi », ou simplement : « Je suis nu ». Dans des versions antérieures des logiciels, les assistants virtuels *firtaient*. Aujourd'hui, certaines des réponses à ces questions ont été remplacées par des phrases standardisées. Par exemple, à la question « Siri, as-tu un copain ? », on peut aujourd'hui entendre : « On m'a dit que mes normes en matière d'intelligence artificielle étaient trop élevées ».

Face à certaines de ces critiques, Siri a réduit la féminisation stéréotypée de son assistant, allant jusqu'à supprimer la voix féminine par défaut en anglais américain (Fournier-Tombs, 2021). Réduire les interactions avec les représentations stéréotypées des femmes peut contribuer à réduire la discrimination et même la violence envers les femmes (Conseil de l'Europe, 2022).

2. DES MÉTHODES ALGORITHMIQUES PEUVENT-ELLES RÉDUIRE LES HYPOTHÈSES DISCRIMINATOIRES ET LES BIAIS SYSTÉMIQUES ENVERS LES FEMMES ?

Les données font partie du problème, mais peuvent aussi faire partie de la solution. Les algorithmes d'apprentissage offrent la possibilité de renverser certaines de ces normes stéréotypées en matière de genre en les entraînant avec des données qui représentent des valeurs culturelles appropriées.

Dans un article de la *Harvard Business Review*, Miller pose la question de manière pertinente : « Comment les biais et les performances des algorithmes se comparent-ils au *statu quo* ? Plutôt que de se demander simplement si les algorithmes sont imparfaits, nous devrions nous demander comment ces défauts se comparent-ils à ceux des êtres humains » (Miller, 2018). Miller examine ensuite plusieurs exemples, tels que les évaluations des candidats à l'emploi ou les évaluations de la récidive, dans lesquels les algorithmes d'apprentissage automatique semblent moins biaisés que les évaluateurs humains.

2.1 Diversité des équipes

Également, il est important de garder à l'esprit que la diversité et l'inclusion dans les équipes de développement de modèles d'intelligence artificielle sont un élément essentiel de la réduction des biais. Comme l'explique Sigal Samuel dans *Vox* : il est difficile d'imaginer une équipe technique composée en majorité de femmes, mettant au point un produit qui répond à la question « Qui est ton papa ? » par « Tu es mon papa » – telle que la réponse donnée par Siri (en anglais : « *Who's your daddy ? You are* ») (Samuel, 2019). Cette anecdote pose la question de qui fait la science des données et pour qui. La place des femmes en technologie, tout comme celle des minorités ethniques et des populations racisées, doit passer de l'utilisation à la production, mais aussi s'appuyer sur l'éducation. L'intégration de plus de femmes dans les études en sciences

et en apprentissage automatique peut contribuer à prendre conscience des problèmes de biais de genre, mais ne suffira pas en soi à les résoudre tous.

2.2 Données biaisées

On peut s'attendre à ce que des équipes de développeurs plus diverses permettent de changer les représentations des femmes dans l'interaction avec les assistants personnels, mais encore faut-il que les femmes scientifiques des données soient aussi conscientes des problèmes liés aux données elles-mêmes (quantité, qualité, données invisibles) et qu'elles soient formées pour remettre en cause les méthodes. Il en est de même, plus globalement, de l'intégration de techniciens des données d'origines ethniques et culturelles variées.

Dans le cas des grands modèles linguistiques, il peut être nécessaire de sélectionner ou de recréer avec soin des données non biaisées, afin de former les modèles. Dans ce sens, l'introduction d'ensembles de données non biaisées dans tous les modèles d'intelligence artificielle permet de lutter contre la discrimination sexuelle accidentelle et intentionnelle dans l'IA.

2.3 Système de valeurs intégré

En outre, encore faut-il que le système de valeurs intégré dans les modèles soit clair et explicite. Bien que les modèles sans paramètres éthiques puissent avoir un effet accélérateur sur l'inégalité entre les genres, les modèles vérifiés peuvent nous aider à mettre en évidence et à inverser nos propres préjugés. Contrairement aux êtres humains, les modèles d'intelligence artificielle n'ont pas d'attachement émotionnel aux valeurs sociales. Dès lors, une intelligence artificielle peut faire mieux que ne pas propager les préjugés sociaux de genre. Elle peut les corriger et diffuser des valeurs universellement reconnues. Certaines de ces valeurs ont été inscrites dans des documents des Nations unies, tels que la *Déclaration universelle des droits de l'homme*⁸ et, surtout, la *Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes* (CEDEF)⁹. Ce texte permettrait de représenter des valeurs universelles envers les femmes à préserver et même à garantir dans les modèles algorithmiques.

8. *Déclaration universelle des droits de l'homme*, Rés AGA 217A (III), Doc off AG NU, 3^e sess, supp n° 13, Doc NU A/810 (1948) 71.

9. *Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes*, 1249 RTNU 13, 18 décembre 1979 (entrée en vigueur : le 3 septembre 1981).

2.4 Lutter contre la discrimination genrée accidentelle dans l’IA

Dans l’ouvrage *Data Feminism* (2020), D’Ignazio et Klein proposent une méthodologie pour lutter contre la discrimination de genre dans les données et la technologie qui intègre les sept principes suivants : (i) examiner le pouvoir, (ii) défier le pouvoir, (iii) susciter les émotions et engagements, (iv) repenser les binaires et les hiérarchies, (v) adopter le pluralisme, (vi) tenir compte du contexte et (vii) rendre le travail visible.

En utilisant ce cadre, nous proposons les solutions suivantes, afin que les assistants vocaux non seulement ne soient pas discriminatoires envers les femmes, mais soient en mesure de renverser la discrimination inhérente à notre société.

i) *Examiner le pouvoir : reconnaître que tous les assistants vocaux les plus courants sont construits par des géants américains du logiciel.*

Les assistants vocaux abordés dans ce chapitre ont tous été conçus par des géants américains du logiciel qui contrôlent une énorme partie du marché. Le tableau ci-dessous montre, par exemple, la taille des entreprises ainsi que le pourcentage d’hommes dans le personnel de programmation.

TABLEAU 20.4
Part du marché et représentation féminine chez les géants américains du logiciel (2018)

NOM DE L’ASSISTANT VIRTUEL	COMPAGNIE	CAPITALISATION BOURSIÈRE ¹⁰	NOMBRE D’EMPLOYÉS MONDIALEMENT	POURCENTAGE DE FEMMES EMPLOYÉES MONDIALEMENT ¹¹
Siri	Apple Inc. (AAPL)	2,126 trillions	147 000	32 %
Alexa	Amazon.com Inc. (AMZN)	1,629 trillion	1 298 000	39 %
Google Home	Alphabet Inc. (GOOG)	1,502 trillion	135 301	31 %
Cortana	Microsoft Inc. (MSFT)	1,878 trillion	96 000	26 %

10. Chiffres obtenus sur Yahoo Finance : <<https://ca.finance.yahoo.com>>.
11. M. Rani, « How Facebook compares to other tech companies in diversity », *Vox*, 2018, <<https://www.vox.com/2018/4/11/17225574/facebook-tech-diversity-women>>.

Cependant, si ces entreprises sont si puissantes, c'est en grande partie parce que l'IA est encore faiblement réglementée. De nombreuses initiatives aujourd'hui en cours viseront, au cours des prochaines années, à protéger les droits humains fondamentaux des utilisateurs. La protection contre la discrimination et les stéréotypes est inscrite dans le droit international des droits de l'homme, notamment dans la *Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes* (CEDEF), qui exige notamment que les pays s'attaquent aux stéréotypes négatifs concernant les femmes dans le secteur privé¹². Cela signifie que l'IA qui propage des stéréotypes négatifs sur les femmes ne sera plus seulement une question d'intérêt économique, puisqu'elle fera l'objet d'une loi sur les droits de l'homme liée à l'IA. Ainsi, la Commission européenne a fait une proposition de règlement sur l'IA le 21 avril 2021¹³ qui imposera des mesures de non-discrimination dans les jeux de données et qui tend à poser un équilibre entre intérêt économique (marché unique numérique de l'Union européenne) et droits fondamentaux.

ii) *Défier le pouvoir : éliminer les biais de genre du financement de la technologie*

Dans ce contexte, lutter contre la discrimination fondée sur le sexe signifie également favoriser un écosystème d'intelligence artificielle diversifié. Les entreprises qui développent des assistants vocaux sont toutes basées aux États-Unis, avec une capitalisation boursière très importante et des dizaines de milliers, des centaines de milliers, voire des millions d'employés. Comme nous l'avons vu, la majorité du financement par capital de risque est également dominée par les hommes, ce qui entraîne une diminution de ce qui est perçu comme des questions féminines.

Au Canada, dans un rapport publié sur le financement de l'IA depuis 2018, Brandusescu constate qu'une grande partie des fonds publics destinés à l'IA vont à un petit nombre d'entreprises détenues par des hommes et à des équipes de recherche dirigées par des hommes (Brandusescu, 2021). Dans ce contexte, remettre en question le pouvoir signifie augmenter la

12. *Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes*, supra note 9.

13. CE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union projet de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle*, dit « Artificial Intelligence Act », COM (2021) 206 final, 21 avril 2021.

pluralité des entreprises ayant accès au financement, non seulement aux États-Unis, mais aussi au Canada.

iii) *Susciter l'émotion et l'incarnation : exiger des tests de sécurité pour tous les utilisateurs (y compris les enfants)*

Les assistants vocaux sont testés pour leur fonctionnalité, mais pas pour leur sécurité. Dans ce contexte, on ignore les dommages, y compris psychologiques, qu'ils peuvent causer envers les enfants, les femmes et d'autres groupes. Leslie, dans un rapport rédigé pour l'Institut Alan Turing, propose un guide pour le développement éthique des systèmes d'IA, qui inclut une référence au « principe de non-préjudice discriminatoire » (Leslie, 2019). Cette idée tourne autour d'une auto-évaluation par le développeur d'une nouvelle technologie, qui peut être rendue plus efficace par une formation appropriée. Si la formation est pertinente, des tests indépendants et obligatoires des technologies d'IA destinées au public seront également nécessaires.

iv) *Repenser les binômes et les hiérarchies : adopter une voix (et une personnalité) non binaire*

Tout comme notre société est en train de repenser les binômes de genre, il est temps de le faire également pour l'intelligence artificielle. Project Q, par exemple, est un assistant vocal non genré, créé par une coalition de chercheurs, dont Copenhagen Pride. « Pensez à moi comme Siri ou Alexa, mais ni homme ni femme », est-il écrit. « Je suis créé pour un avenir où nous ne sommes plus définis par le genre, mais plutôt comment nous nous définissons ». Un autre assistant vocal, Sam, a été développé par Accenture et CereProc. Alors que Q fonctionne principalement en calibrant la hauteur de la voix, Sam intègre également les modèles de discours et les intonations.

v) *Adopter le pluralisme : développer des ensembles de données de formation diversifiés et pluralistes*

Les ensembles de données de formation à l'IA doivent être pluralistes. La formation sur des articles de nouvelles, disponibles sur Google, a perpétué les stéréotypes inhérents à ces articles. De plus, ces articles portent principalement sur l'actualité américaine, alors que les assistants vocaux sont utilisés dans le monde entier. Deux types d'ensembles de données sont nécessaires pour la formation des assistants vocaux : ceux qui sont utilisés

pour l'apprentissage supervisé et ceux qui sont utilisés pour l'apprentissage non supervisé. L'apprentissage supervisé, par exemple un modèle de classification des images, nécessite des données d'entraînement. En contrepartie, l'apprentissage non supervisé regroupe les données semblables (pour créer, par exemple, des profils similaires pour le marketing) et ne nécessite pas de données d'entraînement. Dans le premier cas, les données d'entrée pourraient provenir de sources beaucoup plus diverses, et peut-être même être entraînées par des groupes à parité hommes-femmes.

En ce qui concerne l'apprentissage non supervisé, c'est-à-dire les textes utilisés pour construire des modèles linguistiques permettant aux assistants vocaux de traiter la demande d'un utilisateur et d'y répondre, il est clair qu'une variété de données sources, issues de contextes culturels différents, doit être prise en compte.

Stafanovics et ses collaborateurs proposent une solution pour résoudre les faiblesses des modèles de langage, en particulier ceux qui sont utilisés en traduction, afin d'aller au-delà du fragment de phrase pour trouver des informations contextuelles, afin d'augmenter la précision de la traduction (Stafanovics *et al.*, 2020). Pour aller plus loin dans cette idée, une bonne pratique serait d'exiger que le logiciel de traduction produise un texte neutre en matière de genre, comme *(s)he*, ou *they*, lorsque le contexte de genre n'est pas présent dans le texte source.

vi) ***Prendre en compte le contexte : mesurer l'effet sociétal***

Au-delà des études sur les dommages causés aux individus, comme il est indiqué au point 3, il convient d'examiner les dommages causés à la société par des algorithmes biaisés. La propagation des stéréotypes de genre dans la technologie¹⁴, au-delà des assistants vocaux, peut avoir d'énormes répercussions et contribuer au maintien de la violence domestique et à l'exploitation sexuelle des femmes et des filles. De plus, ces stéréotypes peuvent affecter l'expérience des victimes de violence dans le système de justice, ainsi que l'attitude de la société à leur égard (Gilbert, 2002). De plus, les stéréotypes en lien avec la place de la femme dans le monde du travail, que nous avons observés dans les outils de traduction automatisée, peuvent avoir un effet négatif sur les perceptions des femmes au travail, ainsi que sur leur propre trajectoire. Ainsi, les jeunes filles qui sont exposées à des modèles de langue alliant les femmes au travail de soins plutôt qu'aux

14. Comme les robots sexuels par exemple, dont le « corps » est le plus souvent de genre féminin.

sciences, mathématiques et technologies, auront plus tendance à les éviter (Steinke, 2017). Ces modèles peuvent ainsi inconsciemment influencer les jeunes filles dans leur choix de parcours professionnels en menant à l'autocensure. Plus encore, il a été prouvé que les annonceurs publicitaires d'offres d'emplois proposées en ligne n'envoient pas les mêmes annonces aux hommes et aux femmes, à compétence égale. Les occasions correspondent à des postes moins élevés et à des salaires moindres.

vii) *Rendre le travail visible : atteindre la parité hommes-femmes dans le développement technologique et le leadership*

Ici, nous faisons l'hypothèse que les femmes, ou plutôt les équipes qui sont plus proches de la parité des sexes qu'elles le sont actuellement, sont moins susceptibles de construire des modèles d'IA sexistes que les hommes.

Houser soutient que les outils de recrutement basés sur l'intelligence artificielle, qui sont connus pour être discriminatoires envers les femmes (par exemple, en filtrant les CV contenant le terme « femmes », comme dans « capitaine de l'équipe féminine de crosse »), peuvent être repensés de manière réfléchie afin de réduire la discrimination humaine (Houser, 2019). Elle fait valoir que les humains sont peut-être encore plus discriminatoires que les modèles d'IA, 80 % des dirigeants faisant encore appel à leur instinct et à leurs opinions pour prendre des décisions concernant la gestion des talents. Elle fait plusieurs propositions techniques, telles que l'utilisation de l'IA pour anonymiser les CV, prédire qu'une employée est susceptible de partir et prendre des mesures pour l'empêcher et, surtout, utiliser des ensembles de données non biaisées pour former les modèles en premier lieu.

L'intelligence artificielle est, à la base, une méthodologie de modélisation dont la qualité dépend de celle des données d'entraînement. Donnez au modèle des centaines d'images d'un chat mal étiqueté en tant que chien, et l'outil prédira à tort que les images de chat sont des chiens. Donnez à un modèle de ressources humaines des milliers de CV d'hommes qui ont été embauchés par l'entreprise et le modèle continuera à prédire que les hommes sont les plus susceptibles d'être embauchés. Amazon en a fait l'expérience à ses dépens¹⁵.

15. « Quand le logiciel de recrutement d'Amazon discrimine les femmes », *Les Échos*, 13 octobre 2018, <<https://www.lesechos.fr/industrie-services/conso-distribution/quand-le-logiciel-de-recrutement-damazon-discrimine-les-femmes-141753>>.

CONCLUSION

Comme nous l'avons vu, la discrimination à l'égard des femmes est un problème qui s'est souvent posé lors du développement de nouvelles technologies et l'intelligence artificielle ne fait pas exception. Si la discrimination de genre en intelligence artificielle n'est pas contrôlée, elle aura le potentiel de menacer sérieusement les avancées récentes en matière d'égalité des sexes. L'omniprésence des outils, en particulier auprès des enfants, est aggravée par une combinaison de biais intentionnels et accidentels, qui peuvent être abordés en utilisant un cadre de féminisme des données. Les exemples présentés dans ce chapitre, que ce soit la traduction automatisée, les assistants virtuels, les agents conversationnels, ou même les robots, illustrent un problème plus vaste que l'intelligence artificielle, le plus souvent développée sans considération du genre. Cette réalité d'indifférence (ou fausse neutralité) a pour effet le renforcement des stéréotypes sur les femmes, pouvant mener à de nombreux effets sociaux encore explorés, dont la violence domestique et l'exploitation. En propageant ces normes, l'IA peut aussi les accentuer, en décourageant les filles à poursuivre des carrières en technologie ou en rendant leur embauche plus compliquée.

Ainsi, dans ce chapitre, nous avons proposé sept recommandations qui illustrent l'approche du Data Feminism pour les assistants vocaux, tels que Siri, Alexa, Cortana et bien d'autres. Nous avons choisi de porter une attention particulière aux assistants vocaux, en raison de leur prévalence croissante dans notre société, ainsi que de leur utilisation par les enfants. Cependant, nos recommandations sont aussi applicables dans les autres contextes d'intelligence artificielle que nous avons présentés. Ces propositions ne prétendent pas régler toutes les difficultés liées aux conditions de déploiement de l'IA. Elles doivent naturellement être testées et enrichies en pratique. Cette étude a essentiellement pour vocation de nourrir le dialogue en faveur d'une relance féministe de l'intelligence artificielle¹⁶ !

16. Phrase adaptée du titre de Y. Strengers et J. Kennedy, *The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot*, 2020.

RÉFÉRENCES

Instruments juridiques

Traité internationaux

Convention sur l'élimination de toutes les formes de discrimination à l'égard des femmes, 1249 RTNU 13, 18 décembre 1979 (entrée en vigueur : le 3 septembre 1981).

Déclarations

Déclaration universelle des droits de l'homme, Rés AGA 217A (III), Doc off AG NU, 3^e sess, supp n° 13, Doc NU A/810 (1948) 71.

Législation européenne

CE, *Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union projet de règlement européen établissant des règles harmonisées concernant l'intelligence artificielle*, dit « *Artificial Intelligence Act* », COM (2021) 206 final, 21 avril 2021.

Doctrines

Apple, « Que puis-je demander à Siri ? », 2021, <<https://support.apple.com/en-ca/siri>> (consulté le 1^{er} avril 2021).

Benjamin, Ruha, *Race after Technology: Abolitionist Tools for the New Jim Code*, Polity, 2019.

Bolukbasi, Tolga, et al., « Man is to computer programmer as woman is to homemaker? debiasing word embeddings », *Advances in Neural Information Processing Systems*, n° 29, 2016, p. 4349-4357.

Brandusescu, Ana, *Artificial Intelligence Policy and Funding in Canada: Public Investments, Private Interests*, Centre for Interdisciplinary Research on Montreal, McGill University, mars 2021.

Buolamwini, Joy et Timnit Gebru, « Gender Shades: Intersectional Accuracy Disparities In Commercial Gender Classification », *Proceedings of Machine Learning Research*, vol. 81, p. 1-15, 2018.

Council of Europe, « Preventing gender-based violence. Gender Matters », 2022, <<https://www.coe.int/en/web/gender-matters/prevention-of-gender-based-violence>>.

Crenshaw, Kimberle, « Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics », *University of Chicago Legal Forum*, vol. 1989, n° 1, Art. 8.

Criado Perez, Caroline, *Invisible Women: Data Bias in a World Designed for Men*, Abrams Press, 2019.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee et Kristina Toutanova, « Bert: Pre-training of deep bidirectional transformers for language understanding », arXiv preprint arXiv:1810.04805, 2018.

D'Ignazio Catherine and Klein F. Lauren, *Data Feminism*, The MIT Press, 2020.

Fournier-Tombs, Eleonore, « Apple's Siri is no longer a woman by default, but is this really a win for feminism ? », *The Conversation*, 2021, <<https://theconversation.com/apples-siri-is-no-longer-a-woman-by-default-but-is-this-really-a-win-for-feminism-164030>>.

- Gilbert, Paula R., « Discourses of female violence and societal gender stereotypes », *Violence against Women*, vol. 8, n° 11, 2002, p. 1271-1300.
- King, Hope, « Who is buying Apple products? Old men », *CNN*, 2015, <<https://money.cnn.com/2015/10/29/technology/apple-customers/index.html>> (consulté le 4 juin 2021).
- Kurtz, Annalyn, *Why Secretary is Still the Top Job for Women*, *CNN Business*, 2013 <<https://money.cnn.com/2013/01/31/news/economy/secretary-women-jobs/index.html>> (consulté le 1^{er} avril 2021).
- Houser, Kimberly A., « Can AI solve the diversity problem in the tech industry? Mitigating noise and bias in employment decision-making », *Stan Tech L Rev*, vol. 22, 2019, p. 290-354.
- Leslie, David, « Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector », *The Alan Turing Institute*, 2019.
- Los Feminicidios en Mexico*, « Los Feminicidios en Mexico Data Platform », 2021, <<https://feminicidiosmx.crowdmap.com/main>> (consulté le 1^{er} avril 2021).
- Microsoft, « 2019 Voice report: Consumer adoption of voice technology and digital assistants », 2019, <<https://about.ads.microsoft.com/en-us/insights/2019-voice-report>> (consulté le 1^{er} avril 2021).
- Miller, Alex P., « Want less-biased decisions? Use algorithms », *Harvard Business Review*, 2018, <<https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms>>.
- National Public Media, « The smart audio report », 2020, <<https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/>> (consulté le 1^{er} avril 2021).
- Noble, Safiya U., *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York, NYU Press, 2018.
- Pew Research Center, « Nearly half of Americans use digital voice assistants, mostly on their smartphones », 2017, <<https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/>> (consulté le 1^{er} avril 2021).
- Rani, Molla, « How Facebook compares to other tech companies in diversity », *Vox*, 2018, <<https://www.vox.com/2018/4/11/17225574/facebook-tech-diversity-women>>.
- Samuel, Sigal, « Alexa, are you making me sexist? », *Vox*, 2019, <<https://www.vox.com/future-perfect/2019/6/12/18660353/siri-alexa-sexism-voice-assistants-un-study>>.
- Simonite, Tom, « AI is the future – But where are the women? », *Wired*, 2018, <<https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/>>.
- Stafanovičs Artūrs, Toms Bergmanis et Mārcis Pinnis, « Mitigating gender bias in machine translation with target gender annotations », *ARXIV preprint*, 2020, <<https://arxiv.org/pdf/2010.06203.pdf>>.
- Steele, Jennifer, « Children's gender stereotypes about math: The role of stereotype stratification », *Journal of Applied Social Psychology*, vol. 33, n° 12, 2003, p. 2587-2606.
- Steinke, Jocelyn, « Adolescent girls' STEM identity formation and media images of STEM professionals: Considering the influence of contextual cues », *Frontiers in Psychology*, vol. 8, 2017, p. 716.
- Strengers, Yolande et Jenny Kennedy, *The Smart Wife: Why Siri, Alexa, and Other Smart Home Devices Need a Feminist Reboot*, Cambridge, MIT Press, 2020.

UNESCO et EQUALS Skills Coalition, «I'd blush if I could: Closing gender divides in digital skills through education», *Policy Paper*, 2019.

L'IA DANS LA MODÉRATION

CENSURE DES CONTENUS LGBT SUR TIKTOK ET TWITTER ET TERREAU DE NOUVELLES FORMES SÉMIOTIQUES

Thibault Grison¹ et Virginie Julliard²

Sur TikTok, les défis de beauté ont fait florès auprès des utilisateurs et utilisatrices du réseau social, en particulier depuis le début de la crise sanitaire en 2020 et des périodes de confinement successives. Face caméra, les créateurs et créatrices de contenus se mettent en scène en train de se maquiller tout en donnant des conseils de beauté, livrant des fragments de leur vie personnelle ou suivant simplement le rythme de la chanson attachée à la vidéo. Toutefois, cette tendance a rapidement fait l'objet d'une réappropriation militante par certaines communautés féministes, LGBT ou antiracistes, transposant ces vidéos de beauté en supports privilégiés pour la diffusion de revendications politiques. La visibilité des contenus comme les défis de beauté étant favorisée par les algorithmes de recommandation de TikTok, tenir un discours politique dans une vidéo en apparence anodine, voire d'un type usuellement promu par le site Web, semble être une stratégie efficace pour rendre viral un message politique qui serait habituellement passé sous le joug de la modération automatisée (Cervi, Divon, 2023)³. L'existence de cette pratique suppose donc trois points qui

-
1. Doctorant en sciences de l'information et de la communication à CELSA Sorbonne-Université.
 2. Professeure à l'École des hautes études en sciences de l'information et de la communication (CELSA) de Sorbonne Université.
 3. Esther Newman, «How make-up tutorials became Gen Z's favourite form of TikTok activism», *Dazed Beauty*, 9 février 2021, <<https://www.dazeddigital.com/beauty/head/article/51859/1/tiktok-makeup-tutorials-gen-z-activism-politics-aoc-blm-endsars>>.

guideront notre réflexion tout au long de ce chapitre. D'abord, sur certains réseaux socionumériques (RSN) comme TikTok, les contenus à caractère politique semblent faire l'objet d'un référencement particulier qui tend à les invisibiliser, voire les censurer⁴. Ainsi, l'invisibilisation de certaines publications est due à la fois aux algorithmes de modération qui suppriment certains contenus des RSN, et aux algorithmes de recommandation qui hiérarchisent la visibilité de ceux-ci. Ensuite, du fait de ce processus d'invisibilisation, les internautes sont encouragés à faire preuve d'inventivité pour contourner la modération et tirer profit des mécanismes de mise en visibilité de certaines publications. Enfin, cette inventivité s'illustre notamment par la production de nouvelles formes sémiotiques hybrides qui jouent de la frontière fine entre ce qui est permis sur le RSN et ce qui ne l'est pas, ce qui est valorisé ou non.

Dans ce chapitre, nous étudierons la façon dont ces communautés sexuelles et de genre se mobilisent sur TikTok et Twitter (maintenant X) pour dénoncer la censure dont elles font les frais dans les RSN et les stratégies sémiotiques qu'elles mettent en place pour se rendre visibles en dépit des mécanismes de modération automatisée et de recommandation. Pour cela, nous travaillerons à partir d'une analyse sémio-discursive de 324 gazouillis (*tweets*) publiés par des comptes militants entre le 11 mai 2019 et le 19 octobre 2020 et de 593 vidéos Tiktok publiées entre 2019 et 2021.

Dans un premier temps, nous expliquerons comment le recours à l'IA dans la modération des publications dans les RSN peut paradoxalement engendrer des discriminations algorithmiques contre certaines « catégories protégées ». Nous nous appuierons sur un état de l'art réalisé à partir de travaux anglo-saxons en études culturelles et en études des médias ainsi que d'études et rapports portant sur la régulation de l'IA par les institutions européennes pour produire une typologie des biais algorithmiques intervenant lors de la conception d'un algorithme. Nous expliquerons également comment ces biais peuvent entraîner une invisibilisation de certains contenus en ligne, en particulier de ceux qui sont produits par des communautés sexuelles et de genre.

Dans un deuxième temps, nous mettrons au jour les pratiques d'écriture par lesquelles les communautés sexuelles et de genre silenciées

4. Elena Botella, « TikTok admits it suppressed videos by disabled, queer, and fat creators », *Slate*, 4 décembre 2019, <<https://slate.com/technology/2019/12/tiktok-disabled-users-videos-suppressed.html>>.

parviennent à déjouer la censure, l'invisibilisation ou la moindre visibilité de leurs contenus et se constituent en communautés politiques.

Dans un troisième temps, nous esquisserons une réflexion sur les conditions de possibilité d'une étude de la discrimination algorithmique en contexte de modération. En effet, comment rendre compte, en tant que chercheur ou chercheuse, des invisibilités produites par ce qu'il est commun de considérer comme un objet opaque, une « boîte noire » ? Comment solliciter et recueillir les « savoirs situés » – marginalisés par les objets que nous étudions –, nous rendre sensibles à ceux-ci et comment les rendre visibles, lisibles ?

Ce travail est l'occasion de déconstruire le régime de visibilité propre aux RSN et d'explorer les contre-visualités qui s'esquissent néanmoins en leur sein.

1. DU BIAIS ALGORITHMIQUE À L'INVISIBILISATION DE COMMUNAUTÉS EN LIGNE

1.1 Les biais algorithmiques : comment naissent les discriminations algorithmiques ?

La discrimination algorithmique peut être définie comme le résultat de l'existence de biais dans la conception et la mise en service d'algorithmes. Il s'agit, en d'autres termes, de l'expérience de la discrimination vécue par les utilisateurs et utilisatrices de ces services automatisés. Cette conception de l'IA nécessairement biaisée s'inscrit dans toute une généalogie de travaux en sciences humaines et sociales, dont le propre est de déconstruire la prétendue neutralité des algorithmes (Cardon, 2015). Un champ de recherche visant à penser le rôle de l'IA dans la reproduction des rapports sociaux de race, de genre ou de classe s'est d'ailleurs développé en études des médias et études culturelles au sein de l'Université de Californie à Los Angeles (UCLA), plus spécifiquement du Center for Critical Internet Inquiry (C2i2)⁵. Selon ces approches, l'IA ne serait qu'un ensemble d'opinions matérialisé dans un système mathématique (O'Neil, 2016). Quant aux algorithmes, ils seraient avant tout le reflet des objectifs et des idéologies de leurs concepteurs et concepteurices (Jean, 2019).

Nous appelons « biais algorithmique » la transformation d'une observation générale (souvent stéréotypée) ou statistique en une condition

5. Initialement co-dirigé par les chercheuses Safiya Umoja Noble et Sarah T. Roberts.

algorithmique systématique qui conduit, dans certains cas, à des discriminations technologiques contre certaines populations (Jean, 2019). Il existe plusieurs types de biais qui interviennent à différents moments du cycle de vie d'un algorithme. Nous proposons de les répartir en trois grands ensembles (non-exhaustifs) sur le terrain de la modération automatisée des RSN en nous appuyant sur l'exemple du mauvais référencement du mot « lesbienne », très souvent associé à du contenu à caractère pornographique sur Internet⁶ :

- D'abord, les biais algorithmiques peuvent résulter des ressources d'apprentissage qui, sur le Web, dans leur écrasante majorité, associent (ou conduisent à associer) les termes « femmes » et « lesbiennes » à des contenus pornographiques, du fait même de la forte présence de la pornographie en ligne. Dans ce cas, l'existence de la discrimination algorithmique est imputée à la collecte, au tri ou à la qualité des données à partir desquelles on va entraîner les algorithmes d'apprentissage automatique (Zuiderveen Borgesius, 2018).
- Ensuite il y a un impensé des concepteurs – essentiellement masculins et blancs (Noble, 2018) – qui ne voient pas le déséquilibre produit par les corpus d'apprentissage et associent le lesbianisme à une « culture sexuelle », à la différence de l'hétérosexualité. Dans ce cas, les biais algorithmiques résultent des représentations cognitives des concepteurs d'algorithmes (Bertail *et al.*, 2019) au moment de l'écriture du code, par exemple. Il s'agit là de biais de stéréotypes.
- Enfin, souvent, le critère de rendement économique est défini comme l'objectif principal du développement d'un code informatique, au détriment d'une variable plus éthique non discriminante, par exemple. Les algorithmes développés par les entreprises détentrices des RSN sont en effet optimisés dans une perspective d'efficacité et de rentabilité avant tout (O'Neil, 2016). On peut donc faire deux hypothèses dans le cas que nous citons : soit les concepteurs associent explicitement le mot « lesbienne » à du contenu pornographique en tirant profit de sa popularité en ligne⁷, soit c'est l'importance accordée aux critères écono-

6. Pour plus de détails sur cet enjeu autour du référencement du mot lesbienne, nous renvoyons aux actions menées par le collectif français #SEOLesbienne.

7. Marie Turcan, « Pourquoi le mot "lesbienne" sur Google ne renvoie-t-il que vers des sites pornographiques ? », *Numerama*, 5 avril 2019, <<https://www.numerama.com/>

miques (avec des variables comme le flux de trafic) par rapport à d'autres critères non économiques (comme la contrainte de non-discrimination pour les « catégories protégées ») qui favorise l'émergence de contenus lesbophobes dans le référencement naturel. On parlera alors, dans le second cas, d'un biais de variable omise (Žliobaitė et Custers, 2016)⁸.

Pour résumer, il existe plusieurs types de biais algorithmiques qui conduisent à des discriminations algorithmiques. Selon les cas, les causes divergent et il est souvent compliqué de déterminer avec justesse, même en tant que chercheurs et chercheuses, la source d'une discrimination algorithmique et le biais qui l'a engendrée. En cause : le silence des entreprises sur les modalités de conception de leurs algorithmes, qu'elles justifient comme quelque chose relevant de la propriété intellectuelle. L'algorithme serait, d'une certaine manière, la « recette secrète vitale à leur activité » (O'Neil, 2016).

1.2 Les discriminations algorithmiques à l'encontre des communautés sexuelles et de genre dans les RSN

Les entreprises détentrices des RSN ont, ces dernières années, beaucoup investi dans le recours à l'IA et à l'apprentissage automatique pour décupler leurs capacités de repérage et de modération des contenus illícites⁹. Le travail de modération sur les RSN est réalisé à la fois par des humains – internautes usagers de la plateforme (Nakamura, 2015) et travailleurs et travailleuses du clic (Tubaro, Casilli et Coville, 2020) – et de manière automatisée, sans que l'on connaisse avec certitude les modalités d'articulation de ces méthodes (traitement automatique des signalements produits par les internautes, supervision humaine des traitements automatiques, etc.) ainsi que la part de contenus traités exclusivement de manière automatique. Par ailleurs, si le *Règlement général de protection des données personnelles* (RGPD)¹⁰ rend obligatoire la supervision humaine des

politique/478663-pourquoi-le-mot-lesbienne-sur-google-ne-renvoie-t-il-que-vers-des-sites-pornographiques.html>.

8. Pour plus d'éléments sur les biais de variables omises, voir également Squires, 2003.

9. République française, « Créer un cadre français de responsabilisation des réseaux sociaux : agir en France avec une ambition européenne. Rapport de la mission "Régulation des réseaux sociaux – Expérimentation Facebook" », remis au Secrétaire d'État en charge du numérique, mai 2019, <<https://www.numerique.gouv.fr/uploads/rapport-mission-regulation-reseaux-sociaux.pdf>>.

10. Pour plus d'éléments relatifs au RGPD, nous renvoyons aux travaux de Julien Rossi.

traitements automatiques des contenus en ligne, les débats parlementaires autour de la proposition de loi contre les contenus « haineux » sur Internet en France ont fait craindre à un certain nombre de militants et militantes un recours massif aux algorithmes sans que les moyens humains pour superviser ces traitements suivent (Grisson et Julliard, 2021). Or, force est de constater que la délégation de la modération à l'IA va de pair avec la discrimination à l'encontre des communautés sexuelles, de genre et race. Pour ce qui concerne les internautes LGBT et les travailleurs et travailleuses du sexe (TDS)¹¹, les discriminations algorithmiques se manifestent notamment sous la forme d'une censure abusive de contenus licites. Sur TikTok et Twitter, nous avons ainsi pu recenser plusieurs cas de censure et relever comment celle-ci fonctionnait. Les gazouillis et vidéos Tiktok à partir desquels nous avons travaillé témoignent, en effet, de l'existence de plusieurs cas de suppressions de comptes ou bannissements temporaires ainsi que de censure de contenus, notamment sur Facebook et Twitter. Sur TikTok, les internautes signalent être dans l'impossibilité de choisir un nom d'utilisateur comportant le drapeau LGBT en emoji ou certains mots clés qui pourraient relever d'un discours de haine selon les standards de communauté du RSN. Enfin, cette censure prend aussi des formes plus « insidieuses » (Gatewood *et al.*, 2020) en raison du déréférencement de certains contenus sans que leurs producteurs et productrices en soient informés (le contenu n'est pas supprimé de leur profil bien qu'il soit invisible pour les autres internautes). C'est ce qu'on appelle plus communément la censure fantôme (*shadowban*¹²).

Quelle que soit la forme de cette censure, la cause est souvent la même : l'algorithme ou le modérateur semble assimiler les contenus explicitement LGBT et TDS tantôt à des propos haineux, tantôt à des contenus à caractère pornographique (Grisson et Julliard, 2021). Au premier titre, il convient de rappeler que la militance LGBT et TDS s'est constituée au travers,

11. Bien qu'il convienne de traiter la question des discriminations dans une approche intersectionnelle (notamment en raison des similitudes dans les expériences vécues par les victimes de la modération algorithmique), pour plus d'éléments portant sur la censure abusive vécue par les personnes racisées en ligne, nous renvoyons aux travaux de Safiya Umoja Noble. Pour cette étude nous nous sommes concentrés sur la censure de contenus LGBT comme cas d'école de la discrimination algorithmique dans le contexte de la modération des RSN. Toutefois, du fait de la grande connivence des associations LGBT et TDS dans les luttes militantes, nous nous autorisons à faire des ponts entre ces deux communautés, toutes deux stigmatisées en raison de leur activité sexuelle.

12. Aurore, Gayte, « Qu'est-ce que le shadow ban sur Instagram, et comment s'en sortir ? », *Numerama*, 19 juin 2022, <<https://www.numerama.com/politique/696740-quest-ce-que-le-shadow-ban-sur-instagram.html>>.

notamment, de réappropriation d'insultes (« gouine », « pédé », « queer », « pute »)¹³, or les algorithmes ne parviennent pas à distinguer entre les différents usages possibles de ces mots (visée stigmatisante, affirmation identitaire, etc.). Les publications des militants et militantes comportant un de ces mots sont alors généralement jugées « non conformes aux standards de la communauté » et supprimées. Au second titre, il convient de rappeler que les personnes LGBT sont constamment ramenées à leurs pratiques et cultures sexuelles, lesquelles sont généralement reléguées aux marges de l'espace public (Berlant et Warner, 2018) et restent largement exclues du domaine du visible et du dicible. Il en résulte que les personnes LGBT sont, dans ces nouveaux espaces de publicité, encore invisibilisées, selon des modalités propres aux RSN.

1.3 Modération automatique, régimes de visibilité et (contre)visualités

Chaque RSN peut être considéré comme un *dispositif* d'écriture numérique, susceptible d'articuler des régimes de visibilité, des régimes d'énoncé, des rapports de pouvoir et des lignes de fuite créatives¹⁴. Il repose sur des outils d'écriture, des normes d'interaction et des coûts d'entrée différents (Facebook se distingue par son option familialiste qui le conduit à faire la chasse aux « contenus pornographiques ») susceptibles d'ouvrir des espaces où de nouveaux régimes de visibilité peuvent advenir, notamment pour ce qui concerne les sexualités ou le genre (Julliard, 2022). Les RSN renouvellent les manières de rendre perceptibles les cultures sexuelles et la « différence des sexes », donc de leur conférer une existence. Ils reconduisent des hiérarchies et reproduisent des invisibilisations. L'évolution des techniques visuelles est consubstantielle à l'évolution des régimes de visibilité (Riboni et Bertho, 2020). Toutefois, la notion de visualité, qui renvoie à « la construction culturelle des manières de voir » (Boidy et Martinez Tagliavia, 2018 ; Mitchell, 2014), invite à ne pas se restreindre au registre iconique et nous défendons l'idée selon laquelle les algorithmes utilisés pour la modération et la recommandation participent de ces régimes de visibilité. Outre les objectifs qui leur sont fixés, les algorithmes qui interviennent dans la modération sont plus ou moins efficaces dans le

13. AIDES, « La loi AVIA ne nous rendra pas moins militants-es ! AIDES s'indigne de la censure subie par les activistes LGBTQI+ et les travailleuses-eurs du sexe », <<https://www.aides.org/communiquer/la-loi-avia-ne-nous-rendra-pas-moins-militants-es-aides-indignation-censure-lgbtqi-tds>>.

14. Pour une opérationnalisation de la lecture que Deleuze fait de la notion foucauldienne de dispositif en sciences de la communication, voir Bonaccorsi et Julliard (2010).

repérage de certains signes. Ainsi, les images de corps dénudés sont plus aisément identifiables automatiquement (et supprimés car ils sont considérés comme pornographiques) que le sexisme ou le harcèlement sexuel exprimé textuellement. Il en résulte, pour ce qui concerne notre terrain, que certaines pratiques illicites ne sont pas repérées par la modération (parce qu'elles sont trop nombreuses ou trop complexes à établir), tandis que des pratiques licites font l'objet de suppression abusive.

Dans le cas qui nous intéresse, les suppressions de contenus licites interviennent du fait de la méconnaissance de la militance LGBT, de son lexique, de ses images et des types de mobilisation auxquels elle a donné lieu, qui sont perçus comme « violents », « amoraux » (pour ce qui concerne certaines performances lors des marches des fiertés) ou « provocateurs » (dans le cas des *kiss in* qui ont vocation à rendre visibles les couples de personnes de même sexe dans l'espace public, par exemple). Les matérialités sémiotiques de ces revendications sont à la fois la condition de possibilité de leur expression et ce qui rend celles-ci difficilement lisibles dans les régimes de visibilité hétéronormatifs. En effet, dès lors que les personnes LGBT sont réduites à des pratiques et des cultures sexuelles, et dans la mesure où pratiques et cultures sexuelles sont reléguées dans la sphère privée (Berlant et Warner, 2018), l'accès à l'espace public est problématique.

Notre terrain nous permet ainsi d'étudier la visibilité qui se rapporte aux questions LGBT dans les RSN. La visibilité, rappelle Nicholas Mirzoeff, est « à la fois le *medium* de transmission et de dissémination de l'autorité [et] un mode de médiation pour celles et ceux sur lesquels cette autorité s'exerce » (Mirzoeff, 2011). Il n'en demeure pas moins que des contre-visibilités sont possibles. Ainsi, nous défendons l'idée selon laquelle les personnes LGBT qui se mobilisent contre la suppression abusive et l'invisibilisation de leurs publications, qui trouvent des parades à ces obstacles au travers d'une grande inventivité sémiotique, esquissent ce qui peut relever d'une contre-visibilité. Non seulement ils désignent et déconstruisent les invisibilités produites par le régime de visibilité des RSN ainsi que les manières de voir les questions qui les concernent, mais ils jouent aussi des contraintes imposées par les dispositifs d'écriture numérique en produisant de nouveaux signes pour se désigner et aborder ces enjeux militants.

2. LA FABRIQUE DE STRATÉGIES CONTRE LA CENSURE ALGORITHMIQUE

2.1 Présentation du corpus : des cas d'invisibilisation aux tentatives de visibilité

En travaillant sur des corpus dont la collecte visait *a priori* à recenser des cas de censure supposément algorithmique observés par les militants et militantes LGBT et les TDS, nous avons pu relever tout un ensemble de contenus élaborés dans le but d'échapper à la modération. Notre terrain s'est donc finalement construit de la façon suivante :

Une première partie de notre corpus consiste en une collecte réalisée au moyen du navigateur de recherche de Twitter¹⁵. Les publications sur lesquelles nous avons travaillé portent notamment sur la censure dont les militants et militantes LGBT et TDS étaient victimes en ligne, et la façon dont ils se sont mobilisés contre celle-ci, en France, à l'occasion des débats parlementaires sur la proposition de loi visant à lutter contre les « contenus haineux » sur Internet. Réalisée entre mars 2019 et octobre 2020¹⁶, cette recherche a permis de recenser 324 gazouillis qui ont ensuite été collectés par la Search API de Twitter. Cette collecte a notamment permis d'enregistrer les médias joints. Le corpus comporte donc, outre les 324 gazouillis, 12 articles de presse et 3 communiqués de presse d'associations LGBT et TDS. L'observation quotidienne des comptes auxquels les gazouillis du corpus renvoient nous a invitées à décaler notre focale et à considérer non plus seulement les contenus susceptibles d'entraîner une modération abusive, mais également les stratégies mises en place par certaines communautés en ligne pour éviter la modération.

15. Les mots clés choisis étaient les suivants : « Loi Avia » + [« LGBT » or « PD » or « pédé » or « gouine »]. Ce choix de mots est dû au repérage des propos faisant l'objet d'une modération considérée comme LGBTphobe par les militants et militantes, mais reconnus comme « haineux » par les outils de modération, à l'instar des termes « pédé » ou « gouine ». Ajouter la formule « Loi AVIA » nous a permis de cibler les contenus à caractère militant, engagés sur les débats portant sur la modération abusive des RSN en France.

16. Cette période correspond au moment où beaucoup de comptes militants LGBT et TDS ont été suspendus sur plusieurs RSN tels que Facebook et Twitter en France. Pour ces militants et militantes, ces suppressions étaient dues à un recours plus massif à la modération algorithmique, dans le contexte de préparation de la loi visant à lutter contre les « contenus haineux » sur Internet. Cette proposition de loi prévoyait en effet l'imposition d'un délai de 24 heures pour les entreprises du Web pour retirer les contenus considérés comme pornographiques ou haineux. Compte tenu de ce court délai, de nombreux acteurs opposés à cette disposition ont fait l'hypothèse d'une délégation plus importante du travail de modération à l'IA, pouvant alors encourager une multiplication de discriminations algorithmiques (Grisson et Julliard, 2021).

Dans ce chapitre, nous nous sommes donc employés à étudier :

- l'effet du recours à l'IA dans la modération des RSN sur la visibilité des thématiques LGBT ;
- la façon dont les personnes LGBT et militants et militantes TDS rendent compte de phénomènes de modération abusive en pointant les éléments sémiotiques repérés par les algorithmes (comme les mots clés, par exemple) ;
- les stratégies employées par les communautés sexuelles et de genre pour rendre plus difficilement repérables (donc supprimables) leurs publications.

De ce point de vue, il nous a semblé pertinent et intéressant d'enrichir ce premier corpus d'une deuxième collecte sur TikTok, dont le recours massif aux algorithmes et la censure abusive sont particulièrement décriés¹⁷. Encore peu exploré par les travaux universitaires francophones en raison de sa contemporanéité et de la complexité de la collecte de données sur ce RSN – contrairement à Twitter, par exemple (Gatewood *et al.*, 2020) –, TikTok nous apparaît un terrain contrastant intéressant pour étudier de nouvelles pratiques de dénonciation et de contournement de ce qu'il convient d'appeler une censure.

Au total, nous avons donc recueilli 593 vidéos Tiktok publiées entre 2019 et 2021 grâce à la recherche par mots clés suivante : [« LGBT » + « community guidelines »]¹⁸.

Ces deux collectes complémentaires réalisées sur deux RSN et temporalités différentes ont néanmoins en commun d'être constituées de contenus méta-discursifs. Cette métadiscursivité est présente de deux manières distinctes : d'abord, il s'agit d'un discours portant sur l'impossibilité de discourir (Grisson, 2020) en raison du risque de censure que l'on court en publiant un propos relatif à la modération en ligne ; ensuite, il s'agit d'un discours portant sur la forme employée pour tenir le discours. Il s'agit donc d'un type de discours caractérisé par l'importance donnée à la forme, dont la matérialité sémiotique constitue le propos central. En somme, l'ensemble

17. Nous formulons cette hypothèse au regard de nos échanges avec plusieurs usagers des RSN et de notre lecture de la presse française et anglo-saxonne dans la phase exploratoire de cette étude.

18. Le choix de ces mots clés est le résultat d'une observation participante menée en tant qu'utilisateur ou utilisatrice du RSN afin de cibler sous quel vocable les militants et militantes abordent le sujet de la discrimination algorithmique dont ils font l'objet.

du corpus a été étudié selon une démarche sémio-discursive attentive aux conditions de possibilité d'un discours, aux stratégies rhétoriques et aux formes sémiotiques par lesquelles elles se déploient.

2.2 Des stratégies bricolées pour contrer les algorithmes de modération

Au sein de notre corpus, nous avons pu repérer plusieurs pratiques élaborées en réaction à la censure algorithmique dont les communautés sexuelles et de genre font l'objet sur les RSN. Nous en proposons ici une typologie en trois catégories distinctes, répondant chacune à un objectif et des stratégies particulières :

2.2.1 *Alerter sur la censure par l'IA*

Un premier type de contenus identifiable sur Twitter et TikTok consiste en la publicisation de cas de censure de publications ou profils d'utilisateurs ou utilisatrices LGBT et TDS. Ces publications sont souvent construites à partir de captures d'écran qui font office de « preuve » d'une censure abusive. Il s'agit de captures d'écran des messages d'erreur signalant la suppression d'un contenu ou d'un compte, accompagnés de menaces de bannissement du compte en cas de récidive. Sur TikTok, ce type de contenu est publié sous la forme suivante : la capture d'écran sert d'arrière-plan à la vidéo et la personne détentrice du compte se trouve au premier plan, face caméra, pour ironiser ou s'indigner de la censure dont elle a injustement fait l'objet. Sur Twitter, ce type de contenu est souvent publié par les personnes qui n'ont pas fait directement l'objet de la censure, mais qui vont alerter leur communauté sur le fait qu'une personne qu'ils suivent a été invisibilisée par les algorithmes. Dans les deux cas, il est intéressant de voir que cette pratique a un double avantage : elle permet de rendre publics des cas de censure algorithmique, tout en contribuant à remettre en circulation la publication qui a pourtant fait l'objet d'une modération.

Tous les gazouillis et toutes les vidéos Tiktok pédagogiques visant à informer de l'existence d'une modération algorithmique inefficace peuvent aussi s'insérer dans cette catégorie de contenus. En effet, au sein de notre corpus, nous avons remarqué que certains comptes Tiktok s'étaient donné pour mission de sensibiliser leur communauté aux règles de modération des RSN. C'est le cas de @stephenamon_ ou @zevulous qui ont fait de ce type de contenu un thème récurrent sur leur compte. @zevulous a par ailleurs été à l'initiative d'une pétition sommant l'entreprise détentrice de TikTok à adopter plus de transparence dans l'application de ses *community*

guidelines sur le RSN¹⁹. Il a également proposé à sa communauté de diffuser et remplir un *Google form* pour recenser tous les cas de censure abusive dont iels auraient pu être témoin sur le RSN²⁰. Ce genre de contenus à visée pédagogique et activiste rejoint des démarches militantes entreprises par les associations militantes LGBT et TDS sur Twitter en France. En effet, dans les deux cas, l'objectif est double : sensibiliser au risque de censure algorithmique et interpeller les entreprises détentrices de ces RSN sur leur manque de transparence et leur responsabilité dans l'invisibilisation des militants et militantes LGBT et des TDS en ligne²¹.

2.2.2 Tester l'efficacité des algorithmes

Un second type de publications vise à « tester » l'efficacité algorithmique afin de produire des connaissances utiles à la mobilisation ou afin de déterminer où se situe la limite entre le visible/dicible et l'invisible/indisible sur chaque RSN. Sur TikTok, nous avons notamment observé la façon dont certains utilisateurs ou certaines utilisatrices LGBT produisent des contenus dans le seul but d'obtenir une réaction de la part des modérateurs et des algorithmes de TikTok. Certaines vidéos sont par exemple accompagnées de légendes comme « *hey TT²², does THIS violate community guidelines ?* » (juin 2021) ; « *nothing was there. i did for community guidelines* » (avril 2021) ou encore, « *day 3 of testing community guidelines* » (février 2020). Sur ces vidéos, on peut voir respectivement deux femmes lesbiennes s'enlacer, puis, pour les deux suivantes, un homme ouvertement gay se mettant torse nu face à la caméra. L'intérêt est donc de tester dans quelle mesure le type de contenu visuel peut faire l'objet d'une modération et, si c'est le cas, selon quelle justification.

Sur Twitter, nous avons pu observer des contenus de mise à l'essai relativement similaires. En effet, certains utilisateurs et certaines utilisatrices recourent volontairement à des termes reconnus comme « injurieux » par

19. *Change.org*, « Call on TikTok to release the content moderation guidelines », <https://www.change.org/p/tiktok-call-on-tiktok-to-release-the-content-moderation-guidelines?utm_content=cl_sharecopy_28284285_en-US%3A3&recruiter=14905296&recruited_by_id=759e5710-c94a-012f-f143-40406f61fb41&utm_source=share_petition&utm_medium=copypink&utm_campaign=psf_combo_share_initial&utm_term=share_petition>.

20. Content Moderation Report Form, <tinyurl.com/tiktokreport>.

21. AIDES, « La loi AVIA ne nous rendra pas moins militants-es ! AIDES s'indigne de la censure subie par les activistes LGBTQI+ et les travailleuses-eurs du sexe », <<https://www.aides.org/communiquer/la-loi-avia-ne-nous-rendra-pas-moins-militants-es-aides-indignation-censure-lgbtqi-tds>>.

22. C'est-à-dire « TikTok ».

les outils de reconnaissance automatique pour observer le délai de réaction des algorithmes. Quelques gazouillis se présentent comme suit : « On va tester la loi Avia en 2 secondes là, ça veut dire quoi ton bordel, PD va » (30 mai 2020). L'idée est donc de mesurer la réactivité des algorithmes dans la modération des contenus haineux et de vérifier, par là même, quel mot clé fait potentiellement l'objet d'une censure automatisée. Au-delà du résultat, ce qui nous intéresse ici est la démarche qui cherche à « comprendre comment fonctionne un [dispositif technique] dont on ne dispose pas des plans de conception ou de montage » (Pailler, 2019, p. 55). Ces démarches présentent l'intérêt de rendre possible la production de savoirs techniques sur la modération des RSN depuis l'expérience des dispositifs.

2.2.3 Éviter la censure abusive

Une pratique a particulièrement retenu notre attention : celle consistant à inventer de nouveaux signes ou de nouvelles combinaisons de signes pour se désigner et désigner des thématiques LGBT et TDS. Dans la mesure où ces nouveaux signes et nouvelles combinaisons de signes doivent rester identifiables auprès du public visé, nous avons entretenu une grande proximité avec les comptes LGBT et TDS de notre corpus (abonnement, consultation régulière des publications, etc.) de manière à les repérer. Toutefois, en raison même de leur caractère novateur et inventif, et de leur objectif de passer sous les radars de la modération algorithmique, il n'est pas évident de tous les référencer.

Une première stratégie que nous avons repérée consiste en l'invention de néologismes pour employer certains mots clés reconnus comme « haineux » ou « illicites » par les algorithmes, mais de manière détournée. Citons quelques exemples : le mot gay devient « g4y » ou « gaygaygaygay » ; le terme « pédé » devient « p  » ; « gouine » devient « g0uine » ; « queer » s'écrit « qu33r » ; et « sexe » devient tantôt « seggs », tantôt « S3x3 », etc. Les formes sont tellement infinies qu'il serait vain pour les modérateurs de tenter de toutes les répertorier. Aussi plurielles soient-elles, elles ont la particularité de reprendre les codes du « langage sms », en jouant sémiotiquement entre le langage parlé et le langage écrit et en alternant l'usage des différents claviers – lettres, nombres et émojis étant employés comme s'ils appartenaient au même code linguistique. On appelle cette novlangue l'*algospeak*.

Une seconde stratégie se fonde sur une remédiation des formes de floutage du monde de l'audiovisuel. Dans plusieurs vidéos de notre corpus, les émojis étaient utilisés pour dissimuler les objets, corps ou attitudes

corporelles dont les internautes estiment qu'ils pourraient leur valoir une suppression de leur publication (par exemple, des émojis apposés sur des torsos nus d'hommes gais ou sur un baiser entre deux femmes lesbiennes).

L'emprunt à des formats de vidéos ludiques et populaires sur les RSN constitue la dernière stratégie que nous avons recensée et que les internautes déploient pour éviter de faire l'objet d'une censure abusive. C'est le cas, par exemple, des vidéos de tutoriels de maquillage (que nous citons en introduction) qui comportent un message à visée militante. Ces emprunts à des formats ludiques et populaires permettent non seulement de masquer le discours militant, mais aussi de bénéficier de la faveur des algorithmes de recommandation qui valorisent ce type de vidéo et d'augmenter la portée de leur audience en atteignant de nouveaux publics (*a priori* seulement concernés par les vidéos de maquillage). L'inconvénient est que ces nouveaux publics signalent régulièrement les vidéos qui ne correspondent pas à ce qu'elles prétendent être. Ces vidéos finissent donc par faire l'objet d'une censure. La stratégie que nous évoquons peut donc aussi se retourner contre soi dans un certain nombre de cas.

2.3 La structuration de communautés politiques par la praxis sémiotique

La constitution d'une subjectivité politique, la cohésion du sujet militant procède d'une *praxis* sémiotique : élaboration de vocabulaires et de répertoires de signes communs (Williams, 1985), productions collectives de textes et d'images (Bernheim *et al.*, 2009), grammaires de manifestation (les *kiss in* évoqués ci-dessus), pratiques d'écriture partagées ou appropriations de textes, d'images ou de chansons. Pour le dire autrement, il s'agit moins ici d'observer la façon dont certaines personnes font usage d'un lexique, d'images ou de diverses combinaisons de signes, que de considérer que c'est par des pratiques sémiotiques que des subjectivités et des communautés politiques adviennent (Mitchell, 2014), et plus encore dans les RSN où les communautés ne préexistent généralement pas à leurs pratiques d'écriture et à leurs publications (Julliard, 2022). Selon nous, une communauté politique se structure autour de la contestation des mécanismes d'invisibilisation des contenus LGBT et des pratiques d'écriture nouvelles qui explorent les possibles des dispositifs d'écriture numérique.

Les néologismes pointés ci-dessus sont apparus d'abord dans les publications des minorités sexuelles, de genre et de race pour éviter des vagues de

cyberharcèlement émanant de comptes identitaires et d'extrême droite²³. Dans le contexte d'un recours plus massif à des algorithmes de modération, ces néologismes font l'objet de nouveaux usages et visent à déjouer les mécanismes d'invisibilisation des RSN tels que TikTok. À mesure qu'elles se diffusent, les combinaisons de signes sur lesquelles les néologismes reposent se sont sophistiquées, se chargeant de nouveaux effets de sens. Par exemple, si la graphie « seggs », en lieu et place de « sexe », prétend référer à la même signification, elle n'en élabore pas moins d'autres effets de sens puisqu'elle est également le témoignage de la modération possible de l'autre mot à la place duquel elle se trouve. De même, sur Twitter, l'emploi de « p  » plutôt que « pd » ajoute un sens plus humoristique ou ironique au gazouillis. Progressivement, néologismes, nouvelles combinaisons de signes, floutage et emprunt à des formats évoquant des registres plus récréatifs se sont diffusés dans des sphères sociales plus larges, fédérant une communauté qui dépasse les sphères militantes et associatives.

Le corpus que nous avons collecté sur TikTok comporte, par exemple, une vidéo de @yourassgayfriend, dans laquelle l'auteur interprète un rap en voix hors champ, tandis qu'il apparaît face à la caméra, sa main posée sur sa bouche pour symboliser le fait qu'il est silencieux par les algorithmes. Voici les paroles en question :

*Dear TikTok, it seems you've got nothing to say
and your silence speaks volumes on what
you think is okay.
Racism, homophobia, transphobia and hate...
THIS is the content you let people create.*

*When the good ones speak up, and provide
Truth to Power,
you silence our voices within a few hours.
Your appeals are a joke for the ones
who seek Justice
when your « Community Guidelines » can no
longer be trusted!
But just 'cause you faltered on doing
what's right,*

Cher TikTok, il semble que tu n'as rien à dire
et ton silence en dit long sur ce que tu penses
être acceptable.
Racisme, homophobie, transphobie et haine...
VOILÀ le type de contenu que tu laisses
les gens créer impunément.
Quand les justes s'expriment et disent
la vérité au pouvoir,
tu fais taire nos voix en quelques heures.
Tes propositions de recours sont une
plaisanterie pour ceux qui cherchent la justice
alors que tes « règles de bonne conduite » ne
sont même plus dignes de confiance !
Mais juste parce que tu as failli à faire ce qui
est juste,

23. Cette observation découle de nos échanges avec des militants et militantes antiracistes sur les RSN et de nos expériences en tant qu'utilisateurs ou utilisatrices de ces RSN. À noter que cette observation est également applicable à l'usage que les milieux d'extrême droite font des néologismes pour éviter la censure en ligne et diffuser des discours de haine impunément. Voir notamment Dafaure, 2020.

*doesn't mean we'll all go down this way
without a fight...*

*Raise your fist if you're with me, that TikTok's
got it all wrong...*

*because maybe they'll listen if we speak up...
in song.*

ne veut pas dire que nous suivrons tes traces
comme ça, sans nous battre...

Levez votre poing si vous êtes avec moi,
si vous trouvez que TikTok a tout faux...

parce que peut-être qu'ils écouteront si nous
parlons... en chanson.

Nous voudrions rappeler ici que l'inventivité sémiotique dont les internautes font preuve pour dénoncer et contourner la censure se réalise dans les marges de ce que le dispositif d'écriture numérique rend possible (Julliard, 2018). Les pratiques visant à tirer profit des mécanismes de mise en visibilité de certaines publications par les RSN en témoignent aussi. Par exemple, TikTok offrant la possibilité d'extraire le son d'une vidéo pour recréer du contenu avec une même bande sonore, il en facilite la diffusion (sous une modalité singulière, *a minima* sonore). Cette fonctionnalité de partage a permis que la bande son de la vidéo précitée soit reprise 3430 fois par des tiktokers qui se sont à leur tour filmés face caméra et main devant la bouche pour apporter leur soutien au message particulièrement critique à l'égard de l'entreprise chinoise²⁴, assurant la diffusion de celui-ci.

Il est important de signaler que le repérage de ce type de contenu dans les RSN sur lesquels nous travaillons – pourtant essentiel dans notre réflexion et pour la constitution de notre terrain – est complexe. Nous y sommes parvenus en nous imprégnant des publications des communautés concernées, en suivant quotidiennement ces comptes et leurs abonnés et abonnées pour espérer trouver les contenus qui nous intéressent. Si la recherche par mot clé par le navigateur de recherche du RSN est pratique en ce qu'elle permet de constituer un corpus cohérent sur un même type de contenu, elle ne montre qu'une infime partie des publications concernant le sujet. Nous avancerons deux explications principales à ce phénomène : la première est que les contenus sont, comme nous l'avons expliqué, produits pour ne pas être repérés par des outils de détection automatique ; la seconde est que, par l'entremise du navigateur de recherche du RSN, nous sommes soumis aux résultats que la plateforme a bien voulu nous montrer, sans possibilité de détecter les phénomènes de censure fantôme, particulièrement présents sur notre terrain.

24. TikTok, «Dear TikTok Spoken Word», <https://www.tiktok.com/music/Dear-TikTok-Spoken-Word-6949993320974207750?lang=fr&is_copy_url=1&is_from_webapp=v1>.

3. POUR UNE PRISE EN COMPTE DES SAVOIRS SITUÉS DANS LES ÉTUDES SUR LES DISCRIMINATIONS ALGORITHMIQUES

Étudier ce qui fait l'objet d'une invisibilisation, les processus d'invisibilisation et les pratiques qui cherchent à échapper à ces processus présente un certain nombre de difficultés méthodologiques. Cela implique tout d'abord de cerner ce qui a été rendu invisible, ce que nous avons choisi de faire en recourant aux témoignages des principaux intéressés et principales intéressées, publiés dans les médias et les RSN. Cela implique ensuite de reconstituer le processus d'invisibilisation, dont une partie nous échappe, étant donné l'opacité des algorithmes de modération et de recommandation des entreprises du Web. Nous nous sommes ainsi penchés sur les pratiques de mise à l'essai des internautes concernés, qui ont régulièrement publié des contenus susceptibles d'être modérés, de manière à comprendre le fonctionnement des algorithmes (notamment les signes repérés, et le temps mis à les repérer). Cela implique enfin de repérer des pratiques de contournement qui ont la vocation de ne pas être identifiées comme telles. Pour ce faire, nous avons pris le parti de lire celles et ceux dont les publications pouvaient faire l'objet de modération abusive pour repérer les glissements opérés (néologismes, floutage, réappropriation de formats ludiques et populaires). Lors de nos parcours personnels et de chercheurs et chercheuses, et de notre fréquentation d'objets de recherche voisins, nous avons nous-mêmes éprouvé certains des effets de ces processus d'invisibilisation.

Dans cette dernière section, nous voudrions revenir à la fois sur l'écran que peut constituer la notion d'« opacité » des algorithmes et sur la nécessité d'entrer dans le terrain à partir d'une position située et inspirée des savoirs d'expérience des militants et militantes produits sur l'enjeu de la modération.

3.1 Remettre l'opacité algorithmique à sa juste place

L'opacité des algorithmes semble être le principal obstacle pour les chercheurs et chercheuses dans la production de connaissances sur le fonctionnement de la modération et la recommandation, comme pour le législateur qui souhaite les encadrer. Au premier titre, on soulignera que les nombreux travaux anglo-saxons qui portent sur le sujet adoptent une perspective consistant à appeler à l'ouverture de la « boîte noire » de l'IA (Roberts, 2019 ; O'Neil, 2016). Au second titre, on soulignera que de nombreux rapports plaident pour une plus grande transparence des entreprises quant au fonctionnement de leurs algorithmes (Défenseur

des droits et Commission nationale de l’informatique et des libertés [CNIL], 2020). Entre autres, l’ouverture des données à certaines instances publiques permettrait de conduire des audits issus d’une collaboration entre chercheurs ou chercheuses et institutions publiques sur les algorithmes à fort impact (Bertail *et al.*, 2019). L’opacité de l’IA et son manque de transparence constituent donc un frein épistémologique à la découverte de biais (Défenseur des droits et CNIL, 2020) et à la résolution de discriminations liées au recours à l’IA. Reste qu’il n’est pas toujours possible de rendre transparents et d’expliquer les algorithmes (c’est le cas par exemple des algorithmes d’apprentissage profond qui sont également opaques à leurs concepteurs) et que les algorithmes rendus transparents pourraient être détournés avec des conséquences imprévisibles (Tambe, 2019; Lacroux et Martin-Lacroux, 2021).

Nous estimons que la question de l’opacité ne doit pas faire écran et détourner notre attention. Elle ne doit pas être appréhendée exclusivement comme un obstacle méthodologique, mais également comme un objet inhérent aux enjeux soulevés par la discrimination algorithmique. En effet, on peut considérer l’opacité comme une particularité de la discrimination algorithmique par rapport à d’autres types de discriminations (Gatewood *et al.*, 2020). Contrairement aux autres types de discriminations, il arrive souvent que l’on ne sache pas en être victime. De plus, la prétendue neutralité de l’IA et son fonctionnement par automatisation rendent, de prime abord, l’effet discriminatoire peu crédible. L’opacité constitue donc une entrave au repérage de la discrimination, car elle favorise le doute chez les utilisateurs et utilisatrices victimes de censure injustifiée (Zuiderveen Borgesius, 2018) : « ont-ils *vraiment* été victimes de discrimination ou s’agit-il seulement d’un bug technique ? ».

3.2 Rendre visibles des savoirs marginalisés par les systèmes d’IA

Appréhender le régime de visibilité des RSN en matière d’expérience et de questions LGBT et TDS – au travers des discours de militants et militantes – ne constitue pas uniquement une solution par défaut pour pallier le problème d’opacité des algorithmes de modération et de recommandation. Au contraire, cela constitue une occasion pour saisir les effets concrets de ces dispositifs (Grison *et al.*, 2023).

Les savoirs des personnes victimes des algorithmes et les savoirs militants sont d’autant plus précieux qu’ils sont partie prenante du travail de modération. En effet, comme l’observe Lisa Nakamura, le travail de

modération des RSN est majoritairement réalisé par des personnes issues de « catégories protégées » (telles que les femmes racisées²⁵), sous-payées par des entreprises sous-traitantes. Elle souligne que ce travail est complété par un « travail à risque » en dehors de toute rémunération et réalisé bénévolement par des groupes minorisés présents sur les RSN (Nakamura, 2015). Selon cette conception, la mobilisation militante est partie prenante du travail de modération sur les RSN : tant par les savoirs qui sont produits et diffusés par les activistes que par les méthodes qu'elles mettent en place pour réaliser leur propre travail de modération. En définitive, choisir les discours activistes de communautés sexuelles et de genre comme terrain d'enquête permet de travailler à partir de discours produits « en connaissance de cause » sur ces enjeux-là : elles opèrent des formes de modération tout en étant l'objet de la modération.

Pour conclure, nous insisterons sur le fait que nous appuyer sur les savoirs des personnes concernées et des militants et militantes mobilisés nous permet de contribuer à l'élaboration d'une contre-visualité des questions LGBT – contre-visualité qui expose crûment le dispositif par lequel des contenus licites sont censurés en ligne.

RÉFÉRENCES

- Amabile, Arno, Théophile Lenoir, Tanya Perelmuter et Basile Thodoroff, *Algorithmes : contrôle des biais S.V.P.*, Paris, Institut Montaigne, 2020.
- Angwin, Julia, Jeff Larson, Surya Mattu et Lauren Kirchner, « Machine bias : There's software used across the country to predict future criminals. And it's biased against blacks », *ProPublica*, 2016, <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>.
- Berlant, Lauren, et Michael Warner, « Sexe en public », *Questions de communication*, n° 33, 2018, p. 11-33.
- Bernheim, Cathy, Liliane Kandel, Françoise Picq, Nadja Ringard et al., *mlf//textes premiers*, Paris, Stock, 2009.
- Bertail, Patrice, David Bounie, Stephan Cléménçon et Patrick Waelbroeck, *Algorithmes : biais, discrimination et équité*, Télécom Paris Tech, Fondation Abeona, Institut Mines-Télécom, Institut Carnot, 2019.
- Boidy, Maxime et Francesca Martinez Tagliavia, *Visions et visualités : philosophie politique et culture visuelle*, Poli éditions, 2018.

25. C'est-à-dire les femmes faisant l'expérience de la racisation.

- Bonaccorsi, Julia et Virginie Julliard, « Dispositifs de communication numériques et médiation du politique. Le cas du site Web d'Ideal-Eu », dans M. Aghababae, A. Bonjour, A. Clerc et G. Rauscher (dir.), *Usages et enjeux des dispositifs de médiation*, Nancy, Presses universitaires de Nancy, 2010, p. 65-78.
- Cardon, Dominique, *A quoi rêvent les algorithmes : nos vies à l'heure des big data*, Paris, La République des idées, Seuil, 2015.
- Cervi, Laura, et Tom Divon, « Playful activism : Memetic performances of Palestinian resistance in TikTok #challenges », *Social Media + Society*, n° 1-13, 2023.
- Dafaure, Maxime, « The "great meme war." The alt-right and its multifarious enemies », *Angles*, n° 10, 2020.
- Défenseur des droits, *Algorithmes : prévenir l'automatisation des discriminations*, Commission nationale de l'informatique et des libertés, 2020.
- Gatewood, Cooper et al., *Cartographie de la haine en ligne : tour d'horizon du discours haineux en France*, Institut Supérieur du Droit, 2020.
- Grisson, Thibault, « L'éditorialisation des discours "on peut plus rien dire" : la "liberté d'expression" en fétiche », mémoire de master de recherche, Celsa Sorbonne Université, 2020, <<https://dumas.ccsd.cnrs.fr/dumas-03257859>>.
- Grisson, Thibault et Virginie Julliard, « Les enjeux de la modération automatisée sur les réseaux sociaux numériques : les mobilisations LGBT contre la loi Avia », *Communication, technologies et développement*, n° 10, 2021, <<http://journals.openedition.org/ctd/6049>>.
- Grisson, Thibault, Virginie Julliard, Félix Alié et Victor Ecrement, « La modération abusive sur Twitter. Étude de cas sur l'invisibilisation des contenus LGBT et TDS en ligne », *Réseaux*, vol. 237, n° 1, 2023, p. 119-149.
- Jean, Aurélie, *De l'autre côté de la machine : voyage d'une scientifique au pays des algorithmes*, Paris, Éditions de l'Observatoire, 2019.
- Julliard, Virginie, « La différence des sexes sur Twitter : les conditions d'observabilité d'un engagement affectif et émotionnel », Mémoire d'habilitation à diriger des recherches, Université Paris-Est Créteil, 2018.
- Julliard, Virginie, « Communauté politique, sémiotique, émotionnelle. Ce que la circulation des images révèle de la structuration de la mobilisation anti-genre sur Twitter », *Communication & Langages*, 2022.
- Lambrech, Anja et Catherine Tucker, « Algorithmic Bias ? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads », 2018, <<https://dx.doi.org/10.2139/ssrn.2852260>>.
- Kunert, Stéphanie, « Dégénérer les codes : une pratique sémiotique de défigement », *Semen*, n° 34, 2012.
- Lacroux, Alain et Christelle Martin-Lacroux, « L'intelligence artificielle au service de la lutte contre les discriminations dans le recrutement : nouvelles promesses et nouveaux risques », *Management & Avenir*, n° 122, 2021, p. 121-142.
- Mirzoeff, Nicholas, *The Right to Look : A Counterhistory of Visuality*, Duke University Press, 2011.
- Mitchell, W. J. Thomas, *Que veulent les images ? Une critique de la culture visuelle*, Boidy Maxime, Nicolas Cilins, et Stéphane Roth (trad.), Dijon, Les Presses du réel, 2014.

- Nakamura, Lisa, « The unwanted labour of social media : Women of colour call out culture as venture community management », *New Formations : A Journal of Culture/Theory/Politics*, n° 86, 2015, p. 106-112.
- Noble, Safiya Umoja, *Algorithms of Oppression : How Search Engines Reinforce Racism*, New York University Press, 2018.
- O'Neil, Cathy, *Weapons of Math Destruction : How Big Data Increases Inequality and Threatens Democracy*, New York, Crown, 2016.
- Pailler, Fred, « Les affects classifiés : numérique et médiations sexuelles », thèse de doctorat, Université de Nantes, 2019, <<https://hal.archives-ouvertes.fr/tel-03236631/document>>.
- Riboni, Ulrike Lune et Raphaëlle Bertho, « Introduction au dossier “Les images au cœur des rapports sociaux. Vers de nouveaux régimes de représentation et de visibilité?” », *Études de communication*, n° 54, 2020, p. 7-18.
- Roberts, Sarah T., *Behind the Screen : Content Moderation in the Shadows of Social Media*, New Haven, Yale University Press, 2019.
- Squires, Gregory D., « Racial profiling, insurance style : Insurance redlining and the uneven development of metropolitan areas », *Journal of Urban Affairs*, vol. 25, n° 4, 2003, p. 391-410.
- Tambe, Prasanna, Peter Cappelli et Valery Yakubovich, « Artificial intelligence in human resources management : Challenges and a path forward », *California Management Review*, vol. 61, n° 4, 2019, p. 15-42.
- Tubaro, Paola, Antonio A. Casilli, et Marion Coville, « The trainer, the verifier, the imitator : Three ways in which human platform workers support artificial intelligence », *Big Data & Society*, vol. 7, n° 1, 2020.
- Žliobaitė, Indrė, et Bart Custers, « Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models », *Artificial Intelligence and Law*, vol. 24, n° 2, 2016, p. 183-201.
- Zuiderveen Borgesius, Frederik, *Discrimination, intelligence artificielle et décisions algorithmiques*, Strasbourg, Conseil de l'Europe, 2018.
- Williams, Raymond, *Keywords : A Vocabulary of Culture and Society*, New York, Oxford University Press, 1985.

NOTICES BIOGRAPHIQUES DES CONTRIBUTEURS ET CONTRIBUTRICES

Camille Alegre est juriste Internet Protocol/Information Technology/General Data Protection Regulation (IP/IT/GDPR) chez DSM avocats à la Cour au Luxembourg. Elle est détentrice d'un master of laws (LLM) en droit de la propriété intellectuelle et des technologies de l'information du Trinity College Dublin en Irlande. Elle a précédemment été consultante en protection des données chez Grant Thornton (Luxembourg) entre 2021-2023.

Alexandra Bensamoun est professeure de droit à l'Université Paris-Saclay et spécialiste de droit de la propriété intellectuelle et de droit du numérique. « Personnalité qualifiée » au Conseil supérieur de la propriété littéraire et artistique (CSPLA), du ministère de la Culture français, où elle a conduit plusieurs missions, elle est également experte pour l'UNESCO sur la Convention de 2005 sur la diversité des expressions culturelles. Reconnue pour ses travaux sur le droit de l'intelligence artificielle, elle a été nommée au Comité interministériel de l'intelligence artificielle par la première ministre française en septembre 2023.

Alix Bénistant est maître de conférences en sciences de l'information et de la communication à l'Université Sorbonne Paris Nord et membre du Laboratoire des sciences de l'information et de la communication (LabSIC). Ses recherches portent sur les processus d'industrialisation et de transnationalisation de la culture – en particulier de la musique –, et les enjeux numériques qui leur sont liés. Actuellement, il travaille sur les plateformes culturelles et le rôle des algorithmes de recommandation dans les processus de « découvrabilité » des productions audiovisuelles et musicales.

Christine Blais est candidate au doctorat à l'Université de Montréal. Ses recherches portent sur le marché de l'art et la muséologie à l'ère numérique, notamment la manière dont l'art qui circule sur la chaîne de blocs infléchit les conventions du monde de l'art contemporain. Actuellement responsable du programme éducatif d'initiation au collectionnement SéminArts du Musée d'art contemporain, elle a été chargée de cours en économie de l'art à l'Université de Montréal, a occupé la direction générale de l'Association des galeries d'art contemporain et a œuvré en philanthropie culturelle au Musée des beaux-arts de Montréal.

Olivier Blais est cofondateur et vice-président science de la décision chez Moov AI, un leader québécois en service-conseil en valorisation de données et en intelligence artificielle. Il y dirige l'innovation interne et externe afin de déployer des solutions d'IA

appliquées pour des compagnies comme Pratt & Whitney, Metro, Ricardo et Alloprof. En 2023, il a été nommé président de la délégation canadienne pour l'intelligence artificielle de l'Organisation internationale de normalisation/Commission électrotechnique internationale (ISO/IEC) par le Conseil canadien des normes. Il est également éditeur de la norme ISO mondiale définissant la qualité des systèmes d'IA où il dirige une équipe de 50 experts internationaux de l'IA.

Laure Bolka-Tabary est maîtresse de conférences en sciences de l'information et de la communication à l'Université de Lille et membre du laboratoire du Groupe d'études et de recherche interdisciplinaire en information et communication (Gériico) dont elle anime l'axe « Innovation par l'usage et dispositifs numériques ». À partir d'approches ethnographiques et technosémiotiques, ses travaux portent sur les pratiques numériques ordinaires, les usages des technologies pour la médiation scientifique et muséale et les dynamiques de production, documentarisation, circulation et interprétation des images numériques.

Maude Bonenfant est professeure titulaire au Département de communication sociale et publique de l'Université du Québec à Montréal (UQAM) et docteure en études sémiotiques. Ses recherches sont orientées vers les dimensions sociales des technologies de communication, les réseaux socionumériques, les données massives, l'intelligence artificielle, les communautés en ligne et tout particulièrement l'étude du jeu. Elle est titulaire de la Chaire de recherche du Canada sur les données massives et les communautés de joueurs, codirectrice du Laboratoire de recherche en médias socionumériques et ludification et codirectrice du groupe de recherche Homo Ludens sur le jeu et la communication.

Colette Brin est professeure titulaire au Département d'information et de communication de l'Université Laval, directrice du Centre d'études sur les médias et coresponsable de l'axe Arts, médias et diversité culturelle de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique. Sa recherche et son enseignement s'articulent autour des transformations actuelles du journalisme, notamment par l'entremise des politiques publiques et des initiatives organisationnelles, et des pratiques informationnelles des citoyens. Elle coordonne l'édition canadienne du *Digital News Report* (Reuters Institute for the Study of Journalism) et préside le Comité consultatif indépendant pour l'admissibilité aux mesures fiscales pour soutenir le journalisme de l'Agence du revenu du Canada.

Viêt Cao est le directeur innovation et analytique de données de Synapse C, un organisme sans but lucratif qui aide le milieu culturel à décrire, comprendre et prévoir le monde à partir des données. Il est engagé activement dans des projets de développement de connaissances des publics par la valorisation individuelle ou collective des données de partenaires culturels. Il supervise le développement d'outils de gestion et d'analyse de données et participe à la création de formations à l'attention des acteurs du secteur des arts et de la culture sur les sujets de découvrabilité, gouvernance de données et vitalité culturelle.

Céline Castets-Renard est professeure titulaire à la Faculté de droit civil à l'Université d'Ottawa et titulaire de la chaire de recherche IA responsable à l'échelle mondiale. Elle est titulaire de la chaire de recherche Law, Accountability and social trust au sein de l'Artificial and Natural Intelligence Toulouse Institute (ANITI) et membre honoraire de l'Institut universitaire de France. Elle est affiliée au Yale Internet Society Project et membre experte de l'Observatoire de l'économie des plateformes en ligne à la Commission européenne. Ses recherches portent sur le droit et la régulation de l'intelligence artificielle dans une perspective internationale et comparée.

Philippe Chassignet, archéologue, a fait partie de l'équipe de coordination scientifique du projet Bulliot, Bibracte et moi, de septembre 2019 à juin 2021, comme animateur de communauté, en tant qu'ingénieur d'étude au Centre national de la recherche scientifique (CNRS) rattaché au laboratoire Archéorient (UMR 5133), à la Maison de l'Orient et de la Méditerranée (Lyon).

Christine Chevret-Castellani est enseignante-chercheuse en sciences de l'information et de la communication à l'Université Sorbonne Paris Nord. Elle est spécialisée dans l'analyse des discours d'acteurs publics et industriels concernant la régulation du numérique et de l'intelligence artificielle. Dans le cadre du Laboratoire des sciences de l'information et de la communication (LabSIC) et du Labex industries culturelles et création artistique (Labex ICCA), elle participe à un programme de recherche sur la découvrabilité et à un projet intitulé « Plateformes de SVOD [vidéo à la demande par abonnement] américaines et industries audiovisuelles en Europe. Mutations des marchés et des industries, diversité des contenus ».

Juliette Denis est gestionnaire de projets collaboratifs chez Synapse C. Elle conçoit et structure des réseaux de partenaires culturels et académiques au Québec et au Canada. Facilitatrice, elle anime des activités en collaboration avec ceux-ci pour favoriser l'intelligence collective et les transferts de connaissance. Forte d'une expérience dans la recherche à l'Université Paris 1 Panthéon-Sorbonne et de la publication du livre *La culture des données*, elle met ses compétences en économie de la culture au service des organismes opérant leur virage numérique, en particulier avec les données.

Justine Dorval est candidate au doctorat en communication à l'UQAM, coordonnatrice des activités scientifiques du Centre de recherche interuniversitaire communication, information, société (CRICIS) ainsi que membre du Laboratoire de recherche sur les pratiques audiovisuelles documentaires (Labdoc) de l'UQAM. Elle s'intéresse aux plateformes de diffusion audiovisuelle, au service public et aux pratiques de diffusion du cinéma à partir d'une position ancrée dans l'économie politique de la communication et de la théorie critique. Sa thèse porte sur le rapport de subversion à l'œuvre chez les cinéastes créant une œuvre à l'Office national du film du Canada (ONF) ainsi que dans les départements de la production et de la diffusion numérique.

Lucia Flores Echaiz est avocate en droit criminel et titulaire d'une mineure en philosophie. Elle termine actuellement une maîtrise en droit (LL.M., UQAM) où elle s'intéresse à la discrimination algorithmique et au droit à l'égalité. Son intérêt pour les enjeux éthiques et juridiques entourant l'intelligence artificielle (IA) l'a amenée à corédiger un rapport de l'UNESCO intitulé *Piloter l'IA et les TIC avancées pour les sociétés du savoir* ainsi qu'à travailler comme chercheuse étudiante en éthique de l'IA à l'Algora Lab (UdeM).

Eleonore Fournier-Tombs est directrice de la recherche sur l'action anticipatoire et l'innovation au Centre de recherche politique des Nations unies à New York. Elle est titulaire d'un doctorat en systèmes d'intelligence artificielle délibératifs de l'Université de Genève et d'un postdoctorat de l'Université McGill. En 2023, elle a été désignée par Lighthouse3 comme une des 100 femmes brillantes dans le domaine de l'éthique de l'IA. Ses recherches portent sur la gouvernance de l'IA, les droits des femmes et le multilatéralisme.

Samuel Francblu a eu une première carrière à la Direction des affaires scolaires de la ville de Paris après un master en sciences de l'éducation et de la formation. En 2012, il fonde le Ludolab de L'Assoce, tiers-lieu dédié à la recherche, la création, l'action culturelle et pédagogique, pour y coordonner les démarches de médiation et l'accompagnement des projets ludologiques. En outre doctorant-chercheur en sciences de l'information et de la communication auprès de l'équipe Pixel du Centre de recherche sur les médiations (CREM) de l'Université de Lorraine, il multiplie les articles, communications et participations à différents projets de recherche-action, articulant toujours son travail de recherche avec une pratique de terrain permettant d'envisager les loisirs culturels et la conception de jeu vidéo comme outils de l'ingénierie sociale.

Samuel Gantier est maître de conférences en sciences de l'information et de la communication à l'Université polytechnique Hauts-de-France et membre du Laboratoire de recherche Sociétés et humanités (LARSH). Ses recherches abordent la pratique, la médiation et la médiatisation des films documentaires ainsi que le design d'expérience utilisateur dans le champ des industries culturelles (narrations audiovisuelles interactives et plateformes de vidéo à la demande).

Éric George est professeur titulaire à l'École des médias à l'UQAM, directeur du CRICIS, coéditeur de la revue *Tic&société*, coresponsable du CR 33 Sociologie de la communication et du numérique au sein de l'Association internationale des sociologues de langue française (AISLF) et chercheur associé à la Maison des sciences de l'homme (MSH) de Paris-Nord. Il s'intéresse aux rapports entre communication, capitalisme et démocratie, à partir des perspectives de l'économie politique de la communication, des études culturelles et de la théorie critique. Il a été, entre autres, coordonnateur de deux ouvrages, *Numérisation de la société et enjeux sociopolitiques*, parus en français et en anglais en 2019 et 2020.

Édouard Germain est candidat au doctorat en communication à l'UQAM et membre étudiant du Centre de recherche interuniversitaire sur la communication, l'information et la société

(CRICIS). Dans le cadre de ses recherches, il s'intéresse à la culture populaire et au hip-hop. Son mémoire de maîtrise s'intéressait plus particulièrement aux imaginaires décoloniaux dans les œuvres d'artistes rap provenant des communautés afrodiasporiques au Québec.

Ève Givois est docteure en histoire (Université Paris 1 Panthéon Sorbonne, 2019), chercheuse en histoire et sciences de l'information et de la communication, spécialisée en histoire des médias audiovisuels et médiation des archives audiovisuelles. Elle est aussi gestionnaire de patrimoines audiovisuels (Institut national de l'audiovisuel – INA Sup, 2012). Ce chapitre a été rédigé dans le cadre de son postdoctorat pour le projet AlgoDoc (2020-2022) à l'Université de Lille (laboratoire Gériico) et l'Université polytechnique Hauts-de-France (laboratoire LARSHS). En 2023-2024, elle est chargée de recherche en postdoctorat au Centre Pompidou, Bibliothèque Kandinsky du Musée national d'art moderne – Centre de création industrielle (MNAM-CCI) pour le projet Mémoires audiovisuelles des expositions (MAVEX).

Madelena Gonzalez est professeure de littérature anglophone contemporaine à l'Université d'Avignon où elle dirige l'équipe de recherche pluridisciplinaire d'Identité culturelle, textes et théâtralité (ICTT). Elle a publié un certain nombre de travaux sur la littérature, la culture, l'esthétique et le théâtre, dont trois recueils sur le théâtre des minorités. Elle travaille actuellement sur les rapports entre la science et la littérature et un chapitre intitulé « L'algorithme de la beauté, le jugement esthétique comme science » qui paraîtra en 2024 dans *A Companion to Literary Evaluation* chez Wiley Blackwell.

Thibault Grison est doctorant en sciences de l'information et de la communication rattaché au laboratoire du Groupe de recherches interdisciplinaires sur les processus d'information et de communication (GRIPIC) de l'École des hautes études en sciences de l'information et de la communication (CELSA) Sorbonne-Université. Ses travaux portent sur les enjeux de liberté d'expression et de censure dans les médias. Il réalise actuellement une thèse au sein du Sorbonne Center for Artificial Intelligence (SCAI) sur le risque de discrimination algorithmique qui pèse sur les communautés sexuelles et de genre dans le contexte de la modération des réseaux sociaux numériques. Depuis 2021, il participe aux activités du Centre d'expérimentation en méthodes numériques pour les recherches en sciences humaines et sociales (CERES) avec un intérêt particulier pour l'accessibilité des données des grandes plateformes du numérique pour les chercheurs en sciences sociales.

Véronique Guèvremont est professeure titulaire à la Faculté de droit de l'Université Laval et titulaire de la Chaire UNESCO sur la diversité des expressions culturelles. Elle est coresponsable de l'axe Arts, médias et diversité culturelle de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA). Elle est aussi codirectrice scientifique – Innovations et implications sociétales du Consortium de recherche en intelligence artificielle IVADO. Elle fait partie de la Banque d'experts de l'UNESCO sur la Convention sur la protection et la promotion de la diversité des expressions culturelles et du Groupe d'accompagnement sur la découvrabilité de l'Organisation internationale de la Francophonie (OIF).

Camille Habault était inscrite en thèse à l'Université d'Avignon de 2017 à 2021 et a travaillé sur les robots au théâtre. Elle a participé à plusieurs tables rondes organisées par la Maison Jean-Vilar et fait une intervention sur les robots et l'IA au festival Pint of Science à Avignon. Elle est actuellement en phase de réorientation dans le secteur de l'informatique.

Bernard Jacquemin est maître de conférences en sciences de l'information et de la communication à l'Université de Lille. Il effectue ses recherches au laboratoire Gériico. Ses travaux portent essentiellement sur l'étude des dispositifs de gestion et de traitement de l'information et des données numériques, et des méthodes de description de ces contenus pour optimiser les opérations automatiques sur les contenus. À cet effet, ses recherches incluent la modélisation d'univers culturels et patrimoniaux (ontologies musicales, univers des mines de charbon), la construction collaborative de l'information (Wikipédia) ou l'analyse de l'influence des plateformes sur la diffusion des contenus culturels (films documentaires) ou de la science ouverte (publications, données de la recherche).

Virginie Julliard est professeure en sciences de l'information et de la communication à l'École en hautes études en sciences de l'information et de la communication (CELSA) de Sorbonne Université. Ses travaux portent sur la production médiatique du genre et la constitution de communautés émotionnelles en ligne. Ses recherches actuelles investiguent la structuration du mouvement anti-genre dans les réseaux sociaux numériques. Elle a créé et dirige le Centre d'expérimentation en méthodes numériques pour les recherches en sciences humaines et sociales (CERES) de la Faculté des lettres de Sorbonne Université. En 2012, elle a publié *De la presse à Internet : la parité en question* chez Hermès-Lavoisier.

Eric Kergosien est maître de conférences à l'Université de Lille et membre du laboratoire Gériico. Ses thématiques de recherche sont l'ingénierie documentaire, la fouille de textes et l'ingénierie des connaissances. Il a coordonné plusieurs projets scientifiques d'envergure, dont le projet Tectonik pour la valorisation du patrimoine du textile, a participé activement au projet MEMO MINES sur le patrimoine industriel des mines. Ces projets font intervenir des compétences en modélisation de données hétérogènes, en extraction automatique d'information à partir de corpus textuels et en structuration de base de connaissances ouvertes interopérables. Eric a également travaillé à la mise en place d'un algorithme de recommandation de contenu et à l'évaluation de l'intégration de cet algorithme dans le cadre du projet Algodoc présenté dans ce chapitre.

Alassane Koringho Sagna est doctorant en socioanthropologie du développement et enseignant- vacataire à l'Université Gaston-Berger de Saint-Louis (Sénégal). Son sujet de recherche doctorale porte sur les mobilisations citoyennes dans le cadre de la mise en œuvre de la réforme de l'hydraulique rurale au Sénégal. Au-delà de sa formation en sociologie, il est titulaire d'une licence de journalisme avec une spécialisation sur les questions de développement. Alassane K. Sagna s'intéresse à des thématiques de recherche diverses : sociétés et intelligence artificielle, mobilisations citoyennes, gouvernance des services publics, implémentation des politiques publiques au Sénégal, etc.

Sarah Labelle est professeure des universités en sciences de l'information et de la communication à l'Université Paul-Valéry Montpellier 3 et membre du Laboratoire d'études et de recherches appliquées en sciences sociales (LERASS). Elle mène des recherches sur les transformations médiatiques en cours et les politiques publiques en lien avec le phénomène de « numérisation de la société » : la « société de l'information », « l'administration électronique », « l'ouverture des données ». Sur ces sujets, elle examine des problématiques sur le pouvoir et les médiations tels qu'ils s'exercent et se déploient au sein de dispositifs et d'organisations. Elle mène des enquêtes au sein d'associations, d'institutions et d'administrations publiques pour comprendre le rôle des dispositifs dans la circulation des savoirs, des idées et de la culture tout en prenant en compte les figures du public. Ses recherches dans le champ communicationnel sont ancrées dans une approche sociopolitique, une perspective critique et une démarche anthropologique.

Laura Iseut Lafrance St-Martin est professeure à l'École des arts numériques, de l'animation et du design (NAD) de l'Université du Québec à Chicoutimi, en théorie et pratique de la création de jeux vidéo et docteure en études sémiotiques. Ses recherches portent principalement sur l'expérience des joueurs et joueuses et sur le jeu vidéo comme outil communicationnel, notamment dans l'optique d'un design éthique. Elle s'intéresse aussi aux enjeux épistémologiques et éthiques entourant l'utilisation des données massives et de l'intelligence artificielle. Elle est membre du groupe de recherche Homo Ludens sur le jeu et la communication et chercheuse au sein du laboratoire Praxis (NAD-UQAC).

François Levin est doctorant en philosophie à l'École polytechnique. Ses recherches portent sur les implications épistémologiques et politiques de l'intelligence artificielle. Il a auparavant été rapporteur au Conseil national du numérique et a contribué à la rédaction du rapport Villani visant à établir une stratégie française sur l'intelligence artificielle.

Lily-Cannelle Mathieu (elle) est doctorante en anthropologie socioculturelle à l'Université McGill. Dans sa recherche doctorale, elle propose une approche phénoménologique, sensorielle et féministe aux technologies d'intelligence artificielle (IA) de vision par ordinateur. Lily-Cannelle a obtenu une maîtrise en mobilisation et transfert des connaissances à l'Institut national de la recherche scientifique (INRS) et un baccalauréat honneur bidisciplinaire en anthropologie et histoire de l'art à l'Université McGill. Détentrice des bourses doctorales Vanier-CRSH, Pierre-Elliott-Trudeau (PETF) et Fonds de recherche du Québec, société et culture (FRQSC), Lily-Cannelle a notamment été chercheuse pour la Chaire ABEONA-ENS-OBVIA sur la justice sociale et l'intelligence artificielle de la professeure Karine Gentelet (2020-2022) et pour la Chaire Fernand-Dumont sur la culture de l'INRS.

Serigne Momar Sarr est sociologue et politiste, enseignant-chercheur au Département de sociologie à l'Université Cheikh Anta Diop de Dakar (UCAD) au Sénégal. Ses thèmes de recherches portent sur les communs, l'adaptation au changement climatique, les impacts socioenvironnementaux des industries extractives, les politiques de développement, l'économie sociale et solidaire, l'analyse des politiques publiques, l'administration publique, la jeunesse, l'épistémologie. Il est membre du Laboratoire de sociologie,

anthropologie et psychologie (LASAP/UCAD), de l'International Association for the Study of the Commons (IASC), de l'Association internationale des sociologues de langue française (AISLF) et membre associé au Centre de recherche sur le développement territorial (CRDT-UQAR). En 2020, il a été lauréat de la meilleure thèse de doctorat du Prix de thèses des écoles doctorales du Système universitaire national de l'Académie nationale des sciences et techniques (ANSTS) du Sénégal.

Stephen Monteiro is a media and culture researcher and Assistant Professor of Communication Studies at Concordia University. He is the editor of *The Screen Media Reader* (Bloomsbury) and the author of *The Fabric of Interface* (MIT Press) and *Screen Presence* (Edinburgh University Press).

Emmanuelle Perrin, docteure en histoire et formée aux humanités numériques, a participé à la coordination scientifique du projet Bulliot, Bibracte et moi de septembre 2029 à décembre 2020, dans le cadre des projets HyperThesau et HisArc-RDF, au sein du laboratoire Archéorient (UMR 5133), à la Maison de l'Orient et de la Méditerranée (Lyon).

Iris Richer est doctorante à la Faculté de droit de l'Université Laval. Elle est détentrice d'une maîtrise en droits fondamentaux de l'Université Laval et du diplôme de l'Institut d'études politiques de Rennes en France. Son projet de thèse porte sur la protection du droit de l'enfant de participer à la vie culturelle dans l'environnement numérique et elle est bénéficiaire d'une bourse doctorale de l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA). Elle est membre de la Chaire UNESCO sur la diversité des expressions culturelles et a précédemment travaillé en tant que professionnelle de recherche à la Chaire UNESCO et coordinatrice de l'axe Arts, médias et diversité culturelle de l'OBVIA.

Nicolas St-Germain est diplômé de la maîtrise en communication profil études médiatiques de l'Université du Québec à Montréal (UQAM). Il s'intéresse à l'univers numérique et journalistique, mais plus particulièrement à la relation entre les deux, que ce soient les nouveaux usages que permet la technologie ou le rôle des algorithmes dans la curation des contenus médiatiques. Son mémoire visait justement à comprendre la place des médias locaux dans les résultats de recherche sur Google Actualités.

Lina Uzlyte est chercheuse postdoctorante en sciences de l'information et de la communication associée au Centre de recherche droit et territoire (CRDT), Université de Reims, et affiliée au Centre de recherche sur les liens sociaux (CERLIS). Elle mène une enquête collective au sein d'une équipe pluridisciplinaire sur la définition du métier de guide-conférencier. Sa thèse a été soutenue à l'Université de la Sorbonne nouvelle Paris 3 en 2021, et étudie l'évolution de la compétence des guides en Europe.

Clémence Varin est doctorante en droit international en cotutelle entre l'Université Laval et l'Université de Rennes (France). Ses recherches portent sur la diversité des expressions culturelles dans l'environnement numérique, et plus particulièrement sur la promotion des objectifs et principes de la Convention sur la diversité des expressions culturelles (2005) de l'UNESCO dans d'autres enceintes internationales. Au cours de son parcours doctoral,

elle a obtenu le soutien de plusieurs organismes subventionnaires, dont l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA) et le Fonds de recherche du Québec. Elle est membre de la Chaire UNESCO sur la diversité des expressions culturelles et a précédemment travaillé pendant deux ans au Secrétariat de la Convention de 2005 au siège de l'UNESCO à Paris.

Patrick White est professeur agrégé de journalisme à l'École des médias de l'Université du Québec à Montréal (UQAM). Il s'intéresse aux usages et aux répercussions des technologies en journalisme, comme l'IA et la chaîne de blocs (*blockchain*).

Cet ouvrage examine sous des angles multiples l'influence croissante de l'intelligence artificielle (IA) sur la culture et les médias, explorant à la fois les possibilités et les risques associés à cette technologie. La création artistique, les algorithmes de recommandation pour la diffusion de contenus journalistiques et culturels, la sauvegarde du patrimoine culturel, la protection de la diversité culturelle, les enjeux relatifs à la propriété intellectuelle et le respect des droits fondamentaux font partie des sujets abordés. Si plusieurs textes témoignent d'une certaine confiance placée dans l'IA pour accompagner le créateur, pour augmenter son potentiel, voire pour le protéger, une majorité d'auteurs partagent surtout leurs inquiétudes ou leurs interrogations. L'ouvrage constitue à la fois un guide pour la réflexion et un marqueur pour mesurer l'évolution de ces enjeux dans les années à venir.

VÉRONIQUE GUÈVREMONT est professeure titulaire à la Faculté de droit de l'Université Laval et titulaire de la Chaire UNESCO sur la diversité des expressions culturelles. Ses travaux récents portent notamment sur la découvrabilité des contenus culturels dans l'environnement numérique et la prise en compte des préoccupations culturelles dans les engagements internationaux relatifs au commerce numérique.

COLETTE BRIN est professeure titulaire au Département d'information et de communication de l'Université Laval et directrice du Centre d'études sur les médias. Sa recherche et son enseignement s'articulent autour des transformations actuelles du journalisme, notamment par l'entremise des politiques publiques et des initiatives organisationnelles, et des pratiques informationnelles des citoyens.

Les directrices de cet ouvrage sont coresponsables de l'axe Arts, médias et diversité culturelle de l'Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique (OBVIA).

Illustration de couverture : *Bosphorus*, Refik Anadol Studio

ÉTHIQUE IA ET SOCIÉTÉ

Collection dirigée par Lyse Langlois



OBSERVATOIRE INTERNATIONAL
SUR LES IMPACTS SOCIÉTAUX
DE L'IA ET DU NUMÉRIQUE



Presses de l'Université Laval