



L'intelligence artificielle dans le secteur audiovisuel

IRIS *Spécial*

Une publication
de l'Observatoire européen de l'audiovisuel



IRIS Spécial 2020-2**L'intelligence artificielle dans le secteur audiovisuel**

Observatoire européen de l'audiovisuel, Strasbourg 2020

ISBN 978-92-871-8805-2 (version imprimée)

Directrice de publication – Susanne Nikoltchev, Directrice exécutive

Supervision éditoriale – Maja Cappello, Responsable du département Informations juridiques

Equipe éditoriale – Francisco Javier Cabrera Blázquez, Sophie Valais, Analystes juridiques

Observatoire européen de l'audiovisuel

Auteurs (par ordre alphabétique)

Mira Burri, Sarah Eskens, Kelsey Farish, Giancarlo Frosio, Riccardo Guidotti, Atte Jääskeläinen, Andrea Pin, Justina Raižytė

Traduction

France Courrèges, Julie Mamou, Marco Polo Sarl, Nathalie Sturlèse, Stefan Pooth, Erwin Rohwer, Sonja Schmidt, Ulrike Welsch

Relecture

Catherine Koleda, Gianna Iacino, Anthony Mills

Assistante éditoriale – Sabine Bouajaja

Presse et relations publiques – Alison Hindhaugh, alison.hindhaugh@coe.int

Observatoire européen de l'audiovisuel

Éditeur

Observatoire européen de l'audiovisuel

76, allée de la Robertsau

67000 Strasbourg, France

Tél. : +33 (0)3 90 21 60 00

Fax : +33 (0)3 90 21 60 19

iris.obs@coe.int

www.obs.coe.int

Maquette de couverture – ALTRAN, France

Veillez citer cette publication comme suit :

Cappello M. (Ed.), *IRIS Spécial 2020-2, L'intelligence artificielle dans le secteur audiovisuel*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2020

© Observatoire européen de l'audiovisuel (Conseil de l'Europe), Strasbourg, décembre 2020

Chacune des opinions exprimées dans la publication est personnelle et ne peut en aucun cas être considérée comme représentative du point de vue de l'Observatoire européen de l'audiovisuel, de ses membres ou du Conseil de l'Europe.

L'intelligence artificielle dans le secteur audiovisuel

Mira Burri, Sarah Eskens, Kelsey Farish, Giancarlo Frosio, Riccardo Guidotti, Atte Jääskeläinen, Andrea Pin, Justina Raižytė



Avant-propos

Selon Elon Musk, le fondateur de SpaceX et PDG de Tesla, « nous devons être très prudents avec l'intelligence artificielle », car elle est peut-être « notre plus grande menace existentielle ». Terrifiant, non ? Et pourtant, tout le monde en parle, et de plus en plus d'entreprises y ont recours. C'est l'avenir, ou du moins le dit-on.

Mais qu'est-ce que l'intelligence artificielle ou IA ? Certainement pas la vision dystopique que nous ont dépeinte tant de films hollywoodiens, de *2001, l'odyssée de l'espace* à *Blade Runner* en passant par *Terminator*.

Tout du moins pas encore.

Il est à craindre que la réponse à la question de savoir ce qu'est l'IA soit bien plus prosaïque que cela : en fin de compte, l'IA, ce sont essentiellement des ordinateurs qui font office d'ordinateurs ; des codes logiciels qui englobent des tonnes de données et qui utilisent toutes ces informations brutes selon des instructions prédéfinies.

Bien sûr, son potentiel fascine. La médecine, l'économie, les transports, l'énergie... tous les domaines sont concernés. Ses applications pratiques semblent infinies. Mais comme tout développement technologique, elle n'est pas sans risque. L'IA a aussi un côté obscur, bien sûr, mais probablement pas aussi menaçant qu'Elon Musk voudrait nous le faire croire.

Tout du moins pas encore.

Dans l'industrie audiovisuelle, comme dans d'autres secteurs, le recours croissant à l'IA devrait annoncer un changement de paradigme, car elle peut transformer la chaîne de valeur dans son ensemble : de la production des contenus, la programmation et la publicité, aux attentes et comportements des consommateurs en raison de l'abondance des offres et des dispositifs ainsi que de la personnalisation des contenus. Du côté obscur cependant, l'IA peut contribuer à la prolifération des fausses informations, et elle soulève des questions concernant le droit à l'information, la diversité et le pluralisme des médias, la protection des données, pour n'en citer que quelques-unes.

L'Observatoire européen de l'audiovisuel a décidé d'examiner ces effets de plus près en publiant ce rapport, qui fait suite à un atelier organisé en décembre 2019 sur les opportunités et les défis soulevés par l'IA dans le secteur audiovisuel, en particulier dans le domaine du journalisme et dans le secteur cinématographique. De plus amples informations sur cet événement, y compris un résumé des discussions et des liens vers les présentations des participants, sont disponibles ici :

<https://www.obs.coe.int/fr/web/observatoire/-/workshop-artificial-intelligence-in-the-audiovisual-industry->

Ce rapport, conçu et coordonné par le département juridique de l'Observatoire européen de l'audiovisuel pendant les difficiles mois du confinement, explore différentes

questions nécessitant une analyse d'un point de vue réglementaire et est divisé en deux parties.

La première partie est consacrée aux problèmes d'ordre général et s'ouvre sur le **chapitre 1** rédigé par Riccardo Guidotti, un spécialiste en informatique (Université de Pise). Il explique ce qu'est l'IA et se penche notamment sur le problème des boîtes noires, c'est-à-dire l'absence de transparence quant à la façon dont les systèmes d'IA fonctionnent et prennent des décisions. Il décrit également comment l'IA pourrait être rendue explicable. Les deux chapitres qui suivent cette entrée en matière technique proposent des vues d'ensemble : dans le **chapitre 2**, Andrea Pin (Université de Padoue) explique les problèmes réglementaires qui découlent de la collecte et de l'utilisation de celles qui font tant rêver dans le domaine de l'IA : les données massives. Le **chapitre 3**, de Sarah Eskens (Université d'Amsterdam), propose une vue d'ensemble de questions que pose l'IA en matière de liberté d'expression et d'information, et présente le dispositif juridique qui encadre le recours à l'IA par les médias et l'impact de l'IA sur les droits à la liberté d'expression d'autrui.

La seconde partie de cette publication porte sur des domaines spécifiques du droit et des politiques des médias sur lesquels l'IA pourrait bientôt avoir un impact considérable. La question de la diversité culturelle à l'ère algorithmique est abordée en premier. Mira Burri (Université de Lucerne) explique au **chapitre 4** comment, qu'il s'agisse de la personnalisation de l'information ou des algorithmes de recommandation sur les services de vidéo à la demande, l'IA semble détenir le secret de nos besoins en informations et de nos attentes en matière de divertissement. Elle examine les effets qui s'ensuivent et si ceux-ci justifient une réglementation. Giancarlo Frosio (Centre d'études internationales de la propriété intellectuelle de l'Université de Strasbourg) se penche sur d'autres questions juridiques complexes au **chapitre 5** : si les machines peuvent « créer » des contenus, peuvent-elles détenir des droits d'auteur ? Un individu ou une entreprise peuvent-ils détenir les droits sur une œuvre créée par une machine ? Au **chapitre 6**, Justina Raižytė (Alliance européenne pour l'éthique en publicité) explique comment l'IA ouvre un nouveau champ des possibles pour les annonceurs et, comment elle pourrait être en théorie aussi mieux adaptée au consommateur - non sans soulever néanmoins des questions importantes en matière de respect de la vie privée. Dans le **chapitre 7**, Kelsey Farish (cabinet d'avocats DAC Beachcroft LLP, Londres) nous familiarise avec les questions relevant du droit de la personnalité – les acteurs fantômes, les droits applicables du vivant d'un individu ou après son décès, et les atteintes possibles à ces droits (notamment les *deepfakes*).

Dans la troisième partie de cette publication, Atte Jääskeläinen (Université de LUT et London School of Economics and Political Sciences) présente, au **chapitre 8**, ce que sont, selon lui, les principaux défis réglementaires posés par l'IA dans le secteur audiovisuel, en se concentrant sur les domaines de réglementation possibles ainsi que sur les risques potentiels.

Les textes d'introduction et les remarques finales, rédigés par Francisco Cabrera, analyste juridique senior à l'Observatoire européen de l'audiovisuel, visent à mettre en perspective toutes ces diverses questions juridiques et politiques.

Je souhaite adresser à tous ces brillants auteurs mes plus sincères remerciements pour avoir rendu ce rapport si riche. Et à tous nos lecteurs, je ne peux que souhaiter une bonne lecture !

Strasbourg, décembre 2020

Maja Cappello

Coordinatrice d'IRIS

Responsable du département Informations juridiques

Observatoire européen de l'audiovisuel

Table des matières

1. Intelligence artificielle et explicabilité.....	3
1.1. Qu'est-ce que l'intelligence artificielle ?	3
1.1.1. Une brève histoire de l'intelligence artificielle	4
1.1.2. Des approches différentes de l'intelligence artificielle.....	7
1.1.3. Applications de l'intelligence artificielle	9
1.2. Qu'est-ce que l'intelligence artificielle explicable ?.....	10
1.2.1. Motivations de l'IA.....	12
1.2.2. Les dimensions de l'interprétabilité	13
1.2.3. Comment lire les différentes explications	17
1.3. IA et IA explicable dans le domaine des médias.....	22
1.3.1. Applications d'IA et explicabilité	23
1.3.2. Les services de vidéo à la demande en pratique	27
1.4. Conclusion	28

2. Les données massives ou la matière dont sont faits les rêves de l'IA	31
2.1. Introduction	31
2.2. Le respect de la vie privée comme gardien des données massives.....	33
2.2.1. Les États-Unis d'Amérique.....	33
2.2.2. L'Union européenne.....	34
2.2.3. La Chine.....	36
2.2.4. Trois approches différentes ?.....	37
2.3. Biais et discrimination des données massives	38
2.4. Informer le public : médias, désinformation et contenu illégal	39
2.5. La politique des données massives et la bulle politique	44
2.6. Les médias, chiens de garde de la surveillance ?	46
2.7. Le marché des médias : des stratégies de marché dirigées par les données massives	48
2.8. Approches réglementaires des systèmes basés sur l'IA.....	49
2.9. Conclusion	51

3. Les conséquences du recours à l'intelligence artificielle par les médias d'information sur la liberté d'expression.....	55
3.1. Introduction	55
3.2. Les applications de l'IA au service des médias d'information	56
3.3. Le recours à l'IA par les médias d'information comme élément de la liberté des médias	58

3.3.1. Rôle démocratique des médias d'information	58
3.3.2. Bénéficiaires de la liberté des médias	60
3.3.3. Devoirs et obligations et codes de déontologie journalistique.....	62
3.4. Conséquences de l'IA pour les droits à la liberté d'expression des utilisateurs d'information et des autres participants au débat public	65
3.5. Obligations des États relatives à la liberté des médias.....	68
3.6. Conclusion	70

4. Les politiques de diversité culturelle à l'ère de l'IA 73

4.1. Introduction	73
4.2. Comprendre l'environnement modifié de la création de contenu, de sa distribution, de son utilisation et de sa réutilisation.....	74
4.2.1. Comprendre les nouveaux intermédiaires.....	74
4.2.2. Implications d'agents éditoriaux dirigés par l'IA.....	77
4.3. Pistes d'action possibles : de nouveaux outils pour aborder et impliquer les intermédiaires numériques.....	81
4.3.1. La gouvernance des algorithmes.....	81
4.3.2. La gouvernance par le biais des algorithmes.....	84
4.4. Observations finales	88

5. Droit d'auteur : la machine est-elle un auteur ? 93

5.1. Introduction	93
5.2. Technologie	95
5.3. Protection : la créativité artificielle peut-elle être protégée ?	97
5.3.1. Personnalité : une machine peut-elle être une personne morale ?	97
5.3.2. Paternité : une machine peut-elle être un auteur ?	100
5.3.3. Originalité : une machine peut-elle être originale ?	108
5.4. Options politiques : des incitations sont-elles nécessaires ?	111
5.4.1. Absence de protection : les œuvres créées par IA dans le domaine public	113
5.4.2. Paternité et fictions juridiques : l'auteur doit-il être un humain ?	114
5.4.3. Un robot est-il l'auteur ?	120
5.4.4. Une protection <i>sui generis</i> pour la créativité artificielle	121
5.4.5. Accorder des droits aux éditeurs et aux diffuseurs	122
5.5. Conclusion	122

6. L'intelligence artificielle dans la publicité est-elle en train de devenir un univers sans foi ni loi ou les données sont-elles utilisées à bon escient?127

6.1. Introduction	127
6.2. L'intelligence artificielle dans la publicité : du traçage de la navigation en ligne à la rédaction de scénarios publicitaires.....	129
6.2.1. La publicité programmatique : la bourse aux publicités et aux données.....	129
6.2.2. La créativité algorithmique : l'intelligence artificielle au service de l'imagination.....	133
6.2.3. Des jeux créatifs aux gains	134
6.2.4. Conclusion: la publicité intelligente grâce à l'intelligence artificielle	140
6.3. Les inquiétudes concernant les données massives (Big Data) et l'intelligence artificielle.....	140
6.3.1. Le cadre juridique en vigueur en Europe	142
6.3.2. Conclusion : (essentiellement) le Bon, la Brute et le Truand	144
6.4. L'utilisation de l'intelligence artificielle pour une régulation intelligente de la publicité	145
6.4.1. Les avatars collectent des données à bon escient.....	147
6.4.2. Les progrès de l'intelligence artificielle en matière de conformité publicitaire en France 148	
6.4.3. Tirer parti de la technologie pour renforcer la confiance du marché publicitaire néerlandais.....	150
6.4.4. Les solutions technologiques des géants de l'industrie publicitaire.....	151
6.4.5. La future limite de l'autorégulation de la publicité.....	152
6.5. Conclusion : « La grande ruée vers les données »	153
6.6. Remerciements	155
6.7. Liste des personnes interviewées	156

7. Droits de la personnalité : d'Hollywood aux *deepfakes*.....159

7.1. Introduction	159
7.2. L'IA plante le décor : <i>deepfakes</i> et acteurs fantômes	160
7.2.1. <i>Deepfakes</i>	161
7.2.2. Acteurs fantômes	161
7.3. Les droits de la personnalité et leurs implications	162
7.3.1. Angle 1 : L'image comme propriété (intellectuelle)	163
7.3.2. Angle 2 : Image et reconnaissance de la marque	164
7.3.3. Angle 3 : Protections du droit à la vie privée	165
7.3.4. Angle 4 : La dignité et les droits voisins	166
7.4. Législation dans différents pays.....	168
7.4.1. Allemagne.....	169
7.4.2. France.....	171

7.4.3. Suède.....	172
7.4.4. Guernesey.....	173
7.4.5. Royaume-Uni.....	174
7.4.6. Californie.....	176
7.5. Quel avenir pour le secteur audiovisuel européen ?	177

8. Approches pour un cadre réglementaire durable pour les industries audiovisuelles en Europe183

8.1. Introduction	183
8.1.1. Les bases, simplifiées, de l'IA.....	185
8.2. Comment l'IA est-elle utilisée dans les industries audiovisuelles ?	187
8.3. L'IA est-elle différente des technologies précédentes ?	190
8.3.1. Qui est responsable lorsque l'IA cause un préjudice ?.....	190
8.3.2. Ce n'est pas qu'une question d'économie.....	191
8.4. Nous avons l'obligation morale de faire le bien avec l'IA	192
8.5. La réglementation doit être centrée sur l'homme et fondée sur des objectifs.....	193
8.5.1. Les risques majeurs devraient être pris en compte	194
8.5.2. Les humains sont les responsables.....	195
8.5.3. La transparence comme solution provisoire ?	196
8.6. Centré sur l'humain, et non sur la technologie.....	197

Illustrations

Illustration 1.	Exemple d'explication globale par arborescence décisionnelle obtenue avec TREPAN.....	18
Illustration 2.	Exemple d'explication sous forme de liste de règles obtenue avec CORELS.....	18
Illustration 3.	Exemple d'explication à base de règles factuelles et contrefactuelles obtenue avec LORE	19
Illustration 4.	Exemple d'explication basée sur l'importance des caractéristiques obtenue avec LIME	20
Illustration 5.	Exemple d'explication basée sur l'importance des caractéristiques obtenue avec SHAP	20
Illustration 6.	Exemple de cartes de saillance obtenues par différentes méthodes explicatives. La première colonne montre l'image analysée et le label assigné par le modèle à boîte noire b du système d'IA	21
Illustration 7.	Exemple d'explication par exemples (gauche) et contre-exemples (droite) obtenue avec ABELE. Au-dessus de chaque (contre-)exemple figure le label attribué par le modèle à boîte noire b du système d'IA.	22

Tableaux

Tableau 1.	Glossaire de la publicité programmatique.....	130
Tableau 2.	Campagnes de publicité et de marketing optimisées par les technologies créatives de l'intelligence artificielle	135

Liste des abréviations

AEEP/EASA	Alliance européenne pour l'éthique en publicité (European Advertising Standards Alliance)
AI4SG	IA pour le bien social (AI for social good)
CAN	Réseau antagoniste créatif (<i>creative adversarial network</i>)
CEDH	Convention européenne des droits de l'homme
DSP	Plateforme <i>demand-side</i> (<i>demand-side-platform</i>)
EASA/AEEP	European Advertising Standards Alliance (Alliance européenne pour l'éthique en publicité)
EPRS	Service de recherche du Parlement européen
GAN	Réseaux antagonistes génératifs (<i>generative adversarial networks</i>)
GEE	Groupe européen d'éthique des sciences et des nouvelles technologies
GEP	Guides électroniques des programmes
GPS	General Problem Solver (conçu pour imiter les comportements humains de résolution de problèmes)
GPU	Processeurs graphiques
IA	Intelligence artificielle
IAB	Interactive Advertising Bureau
IAE	Intelligence artificielle explicable
LT	Logic Theorist (premier programme capable de raisonner)
NLG	Génération de langage naturel (<i>natural language generation</i>)
NLP	Traitement naturel du langage (<i>natural language processing</i>)
RGDP	Le Règlement général sur la protection des données
RNA	Réseaux de neurones artificiels
RNP	Réseaux de neurones profonds
RTB	Enchères en temps réel (<i>real-time bidding</i>)
SSP	Plateforme <i>Supply-side</i> (<i>supply-side platform</i>)
VFX	Effets spéciaux visuels (<i>visual special effects</i>)
WMFH	travail sur commande (<i>work-made-for-hire</i>)

La boîte noire

Comme nous le disions dans l'avant-propos de cette publication, l'IA est un phénomène aussi fascinant que terrifiant. Ses succès actuels, tout comme son potentiel, impressionnent, et les différents chapitres de cette publication illustrent bien les diverses manières dont l'IA pourrait révolutionner (ou révolutionne déjà) le secteur audiovisuel. Des machines intelligentes peuvent composer de la musique ou les paroles d'une chanson, vous dire ce que vous devriez regarder ou lire, ou même ramener (virtuellement) des morts à la vie ! C'est peut-être pour ces raisons que l'IA, comme n'importe quelle découverte technologique perturbatrice avant elle, suscite un sentiment de crainte. Cela n'a rien d'étonnant : la peur de ce que l'on ne peut ni comprendre ni contrôler est le lot de la nature humaine. C'est pourquoi le problème le plus urgent à résoudre lorsqu'on s'interroge sur la réglementation de l'IA est celui dit de la boîte noire. Comme l'explique Riccardo Guidotti dans sa contribution à cette publication, « les modèles à boîte noire sont des outils sur lesquels s'appuie l'IA pour accomplir des tâches et dont le processus décisionnel n'est soit pas accessible, soit non compréhensible pour l'homme. » En d'autres termes, il s'agit de machines qui prennent des décisions qui affectent la vie humaine, sans qu'intervienne un contrôle humain, ou sans que les raisons de leurs décisions soient connues des humains. Selon Guidotti, il ne s'agit pas uniquement d'un problème « du problème que posent les possibles biais dont les modèles à boîte noire héritent à partir des artefacts et des préjugés qui sont dissimulés dans les données utilisées pour l'apprentissage automatique des systèmes d'IA obscurs. » Ainsi, l'absence de transparence est l'un des principaux problèmes que soulève aujourd'hui le recours aux algorithmes. Si, comme on l'entend parfois, un algorithme est comme une recette de cuisine, alors les algorithmes utilisés par certaines entreprises sont un peu comme la recette du Coca-Cola, le secret le mieux gardé au monde. Mais il est également vrai que beaucoup de gens adoptent vis-à-vis des algorithmes la même approche qu'avec certains aliments : tant qu'ils apprécient ce qu'ils mangent, peu leur importe la recette, et la plupart du temps, ils préfèrent d'ailleurs ne pas en connaître les ingrédients. Quoi qu'il en soit, du moins dans certains cas extrêmes, il y a de bonnes raisons d'avoir peur. C'est pourquoi les experts appellent à faire en sorte que le fonctionnement et les résultats des systèmes d'IA soient explicables.

1. Intelligence artificielle et explicabilité

Riccardo Guidotti, Université de Pise

L'intelligence artificielle est aujourd'hui l'un des domaines scientifiques et technologiques les plus importants. Son impact socio-économique est considérable, et son adoption se généralise à tous les champs de la société de l'information moderne. Les applications de premier plan basées sur l'intelligence artificielle comprennent les assistants vocaux (par exemple Siri et Alexa), les véhicules autonomes (par exemple, les voitures autonomes, les drones, les robots de nettoyage), le diagnostic médical, le filtrage des spams et la reconnaissance d'images. Les systèmes d'intelligence artificielle connaissent une réussite impressionnante en émulant le comportement humain principalement par le biais de modèles d'apprentissage automatique (*machine learning*) obscurs. Ces modèles sont généralement fondés sur des réseaux de neurones profonds (*deep neuronal networks*), qui ne permettent pas de connaître la logique selon laquelle leurs processus internes fonctionnent.

L'absence de transparence du processus décisionnel suivi par ces modèles pose une question éthique fondamentale, et limite leur adoption dans des domaines critiques du point de vue de la sécurité ou qui sont socialement sensibles. En effet, il ne s'agit pas uniquement d'absence de transparence, mais aussi du problème que posent les possibles biais dont les modèles à boîte noire héritent à partir des artefacts et des préjugés qui sont dissimulés dans les données utilisées pour l'apprentissage automatique des systèmes d'IA obscurs. De plus, l'intelligence artificielle peut être utilisée pour créer des contenus artificiels réalistes. L'intelligence artificielle et les possibilités de monétisation qu'elle engendre changent de manière profonde les industries des médias et du divertissement, des recommandations personnalisées à la création de contenus.

1.1. Qu'est-ce que l'intelligence artificielle ?

L'intelligence artificielle (IA) est l'« intelligence » dont font preuve les machines ou tout autre logiciel ou technologie lorsqu'ils effectuent une tâche¹. Le terme « artificiel » est employé pour la distinguer de l'intelligence « naturelle » ou « biologique » des humains. L'IA est un domaine de recherche en informatique qui tente de comprendre l'essence de l'intelligence et de créer des machines intelligentes qui raisonnent et répondent en simulant l'intelligence humaine. L'étude de l'IA est traditionnellement considérée comme

¹ Russell, S. et Norvig, P., *Artificial intelligence : a modern approach*, Pearson.

l'étude d'« agents intelligents » qui perçoivent un *environnement* et qui exécutent des *actions* pour augmenter leurs chances d'atteindre un objectif préalablement défini². Depuis la naissance de l'IA, les théories et technologies qui la sous-tendent sont devenues de plus en plus matures, et ses champs d'application se sont élargis.

1.1.1. Une brève histoire de l'intelligence artificielle

Le terme d'« intelligence artificielle » a été proposé en 1956 par John McCarthy au cours d'un atelier à l'Université de Dartmouth³ pour distinguer l'IA de la cybernétique⁴. Il est admis que cet atelier marque la naissance de l'IA.

1.1.1.1. Les premières années de l'IA

Les premières années de l'IA (1952 – 1969) ont été marquées par de nombreux succès, limités par la capacité des ordinateurs élémentaires de l'époque qui n'étaient considérés que comme de puissantes calculatrices, uniquement à même de faire des mathématiques. Après « Logic Theorist » (LT), le premier programme capable de raisonner, Allen Newell et Herbert Simon ont créé le « General Problem Solver » (GPS), qui, à la différence de LT, avait été conçu pour imiter les comportements humains de résolution de problèmes. Grâce au GPS, Newell et Simon ont pu formuler la célèbre *hypothèse du système de symbole physique*, selon laquelle tout système intelligent fonctionne en manipulant des symboles. En 1958, John McCarthy du MIT a mis au point le langage de programmation de haut niveau Lisp, qui a été jusqu'aux années 1990 le langage dominant de l'IA⁵. Dans les années 1960, de nombreuses pistes de recherche⁶ fructueuses ont été suivies dans le domaine de l'IA.

1.1.1.2. Le premier hiver de l'IA

De 1970 à 1980, l'IA a connu son premier « hiver », au cours duquel la recherche a connu un ralentissement constant⁷. Les promesses de progrès des chercheurs en IA n'ont pas été tenues en raison des limites techniques imposées par les ordinateurs utilisés pour mettre en œuvre les programmes d'IA, qui n'étaient capables de résoudre que des « jeux »⁸. Ils ne

² Poole, D., Mackworth, A., et Goebel, R., *Computational Intelligence*, Pearson.

³ Crevier, D. (1993), *AI : the tumultuous history of the search for artificial intelligence*, Basic Books, Inc.

⁴ L'IA et la cybernétique sont deux champs d'études différents mais connectés qui se fondent sur le même principe de logique binaire. Cependant, l'IA cherche à créer des machines qui imitent l'intelligence humaine et peuvent se comporter comme des humains, tandis que la cybernétique se base sur une vision constructiviste du monde, et se concentre sur les interactions entre humains et machines : la manière dont un système traite l'information, y répond et s'y adapte. Les différences entre IA et cybernétique ne sont donc pas simplement sémantiques mais aussi conceptuelles.

⁵ Reilly, E. D., *Milestones in computer science and information technology*, Greenwood Publishing Group.

⁶ McCorduck, P. et Cfe, C., *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*, CRC Press.

⁷ Russell, S. et Norvig, P., *op.cit.*

⁸ Crevier, D., *op.cit.*

disposaient pas encore de suffisamment de mémoire ou de la vitesse de traitement nécessaire pour faire quoi que ce soit de réellement utile. Les systèmes d'IA logiques introduits par McCarthy, qui exécutaient des programmes de déduction, n'étaient pas capables de résoudre des problèmes réels, car ils exigeaient un nombre considérable d'étapes pour démontrer des théorèmes très simples⁹. De plus, de nombreux programmes d'IA ont en pratique besoin de quantités gigantesques de données. Malheureusement, à l'époque, personne n'avait réuni ou été à même de réunir des bases de données suffisamment grandes. Les années 1970 ont vu la naissance de Prolog¹⁰, langage de programmation logique performant, qui appliquait une logique plus efficace d'IA permettant des calculs complexes. Les détracteurs de l'approche logique ont lancé un débat entre, d'une part, ceux qui soutenaient que les machines devaient penser comme les humains, et ceux pour lesquels les machines devaient résoudre les problèmes de façon indépendante de la manière dont le font les humains. Les agences qui finançaient la recherche en IA ont ainsi été déçues en raison de l'absence de progrès réalisés et ont coupé presque tous les fonds alloués à la recherche dans ce domaine. Au même moment, la recherche sur les réseaux de neurones (*neural networks*) fut interrompue pendant près de dix ans à la suite de la parution du livre *Perceptrons* en 1969. Un perceptron¹¹ est une forme primitive de réseau de neurones, et de nos jours, les réseaux de neurones constituent un élément fondamental des systèmes d'IA modernes. Un réseau de neurones (artificiel) est un modèle d'apprentissage automatique inspiré des modèles neuronaux biologiques et composé de neurones artificiels. Il reçoit une entrée, l'associe à l'état interne des neurones, et produit une sortie en utilisant une fonction d'activation. Les entrées sont des données, par exemple des tableaux, images ou documents, et la sortie consiste en une classification. Un réseau de neurones apprend à produire une sortie sur la base d'une entrée donnée à partir d'un ensemble de données d'apprentissage annotées.

1.1.1.3. Le boom de l'IA

Dans les années 1980, le « savoir » a été placé au cœur de la recherche en IA, et de nombreuses entreprises ont commencé à adopter des formes d'IA appelées des « systèmes expert » (*expert system*). Un système expert est un algorithme qui, en exploitant un savoir donné représenté par des règles de type « si-alors » (*if-then*), imite la capacité décisionnelle d'un expert humain¹². Un système expert est formé d'une « base de connaissances » (*knowledge base*) qui représente des faits et des règles et d'un « moteur d'inférence » (*inference engine*), qui applique les règles aux faits connus pour en déduire de nouveaux faits. Les systèmes expert ont fait partie des premiers programmes logiciels d'IA efficaces à être adoptés par certaines entreprises. Les chercheurs ont réalisé que la puissance des systèmes expert venait des connaissances qu'ils renfermaient et que « ... l'intelligence pourrait être basée sur la capacité à utiliser une large quantité de connaissances diverses de différentes manières¹³ ». Cette injection de confiance dans l'IA a poussé les financiers à

⁹ McCorduck, P. et Cfe, C., *op.cit.*

¹⁰ Crevier, D., *op.cit.*

¹¹ Tan, P.-N. *et al.*, *Introduction to data mining*, Pearson Education India.

¹² Jackson, P., *Introduction to expert systems*, Addison-Wesley Longman Publishing Co., Inc.

¹³ McCorduck, P. et Cfe, C., *op.cit.*

investir de nouveau dans la recherche en IA. En parallèle, les réseaux de neurones ont connu une « renaissance ». Hinton et Rumelhard ont popularisé la « rétropropagation¹⁴ » (*backpropagation*), une méthode efficace d'apprentissage pour les réseaux de neurones. Cette méthode d'apprentissage a rendu efficace le recours aux réseaux de neurones artificiels (RNA) (*artificial neural networks*), des systèmes d'apprentissage automatique inspirés des réseaux de neurones biologiques du cerveau humain¹⁵. Les RNA « apprennent » à partir des exemples fournis par un ensemble de données comment évaluer une tâche sans qu'il leur soit nécessaire de disposer de règles préexistantes spécifiques à cette tâche. Par exemple, ils peuvent reconnaître si une image représente un piéton ou une voiture à partir d'images dont le contenu a été légendé et sans connaissance préalable des objets étudiés.

1.1.1.4. Le deuxième hiver de l'IA

Le financement de la recherche en IA a été inégal pendant le « deuxième hiver de l'IA » (1987 – 1993). La puissance et la vitesse des ordinateurs de bureau d'Apple et d'IBM, qui ne reposaient sur aucune forme d'IA, s'amélioraient lentement, et en 1987, ils devinrent plus efficaces que les machines coûteuses Prolog et Lisp. Néanmoins, malgré les critiques de certains investisseurs et gouvernements, l'IA a continué d'avancer. Au cours de ces années, le concept d'« agents intelligents » a été finalisé, grâce à la définition en économie de « l'agent rationnel ». Un agent intelligent est un système qui effectue des actions qui maximisent les chances de succès vis-à-vis d'un objectif prédéfini. En outre, l'IA devint une discipline scientifique « rigoureuse » parce que les chercheurs en IA ont de plus en plus recouru à des outils mathématiques sophistiqués pour développer les programmes d'IA. Par exemple, les probabilités et la théorie de la décision ont été amenées dans le champ de l'IA avec le livre de Judea Pearl¹⁶. Cependant, malgré ces progrès évidents, l'IA en tant que champ de recherche théorique ne retenait pas encore véritablement l'attention, dans la mesure où les algorithmes élaborés initialement pour l'IA commençaient d'être exploités dans le cadre de systèmes plus larges de l'industrie technologique, par exemple l'exploration de données, le diagnostic médical, la reconnaissance vocale, les moteurs de recherche, les logiciels bancaires, la robotique industrielle, etc.

1.1.1.5. Données massives, apprentissage profond et IA

En dépit des avancées mentionnées plus haut, le tournant décisif de l'IA a surtout résulté de l'augmentation colossale de la puissance des ordinateurs dans les années 1990. Deep Blue¹⁷ et Watson¹⁸ sont des exemples très célèbres de succès en IA qui découlent de ces avancées technologiques. Le Deep Blue d'IBM a été le premier système d'IA de jeu d'échecs à battre un champion du monde, Garry Kasparov¹⁹, en 1997. En 2011, le système Watson

¹⁴ Rumelhart, D., Hinton, G. & Williams, R., « Learning representations by back-propagating errors », *Nature* 323, 533–536 (1986), <https://doi.org/10.1038/323533a0>.

¹⁵ Tan, P.-N. *et al.*, *op.cit.*

¹⁶ Pearl, J. (1988), *Probabilistic reasoning in intelligent systems*.

¹⁷ Disponible sur : <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/>.

¹⁸ Disponible sur : <https://www.ibm.com/ibm/history/ibm100/us/en/icons/watson/>.

¹⁹ Russell, S. et Norvig, P., *op.cit.*

d'IBM, conçu pour répondre à des questions, a battu avec une marge considérable les champions du jeu télévisé « Jeopardy ! ». En outre, à partir de 2010, en plus des avancées en puissance informatique, l'IA est entrée dans une nouvelle ère en raison des progrès technologiques réalisés en termes de capacités de stockage, de facilité d'accès aux données massives (*big data*), et de l'avancée des techniques d'apprentissage automatique telle que les réseaux de neurones profonds (*deep neural networks*).

- Le terme de « données massives » désigne une quantité de données considérable qui ne peut être stockée, gérée et traitée par un logiciel conventionnel. L'ère des données massives est née de deux principaux flux :
 - (i) les secteurs industriels qui stockent de l'information : des registres d'activités aux achats des clients ;
 - (ii) la collecte généralisée par les smartphones et les équipements mobiles d'informations personnelles relatives aux utilisateurs à partir de sources variées telles que les messages postés sur les réseaux sociaux, les courriers électroniques, les traces de mobilité, les dossiers médicaux, etc.
- Les réseaux de neurones profonds (RNP) sont des modèles conçus comme une évolution des réseaux de neurones artificiels (RNA) et qui se composent de nombreuses couches de traitement (profondeur). L'apprentissage profond est une branche de l'apprentissage automatique qui étudie les RNP. Les RNP peuvent être appliqués au traitement de tâches bien plus complexes que celles qui peuvent être résolues par les RNA, par exemple la reconnaissance d'images, la reconnaissance vocale, le traitement du langage naturel, etc. Mais la popularité récente des RNP est avant tout due aux nouveaux processeurs graphiques (GPU). Ceux-ci ont entraîné une accélération marquée des processus d'apprentissage des RNP et ont renforcé leur efficacité par rapport aux processeurs traditionnels (les unités centrales de traitement). Malheureusement, comme nous le verrons dans le prochain chapitre, les RNP ont un inconvénient fondamental : la difficulté à être interprétés²⁰.

1.1.2. Des approches différentes de l'intelligence artificielle

On retient habituellement quatre notions différentes pour aborder l'IA²¹, dans le cadre de deux dimensions :

1. Observer la manière artificielle de *penser* ou le *comportement* artificiel ;
2. Modéliser des *humains* ou modéliser une norme *idéale* (appelée rationalité).

Les quatre notions différentes sont donc la « pensée humaine », la « pensée rationnelle », le « comportement humain » et le « comportement rationnel ». Divers chercheurs, avec des approches différentes, se sont alignés sur ces quatre notions. Ainsi, la recherche portant sur

²⁰ Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., et Pedreschi, D. (2018), « A survey of methods for explaining black box models », *ACM computing surveys (CSUR)*, 51(5):1–42.

²¹ Russell, S. et Norvig, P., *op.cit.*

l'IA a été divisée en sous-domaines qui peinent souvent à communiquer entre eux. Ces sous-domaines peuvent être différenciés en termes de notions et de variantes philosophiques, en ce qui concerne les objectifs particuliers qu'ils cherchent à atteindre et les méthodes techniques qu'ils emploient.

S'agissant des différences philosophiques, l'approche centrée sur l'humain et l'approche rationnelle se distinguent. Selon l'approche centrée sur l'humain, l'IA devrait simuler une intelligence naturelle. À l'inverse, l'approche rationnelle, qui implique une combinaison de mathématiques et d'ingénierie, avance que la biologie humaine est hors propos. Selon ce point de vue, l'IA peut être soit développée par des principes simples et élégants, tels que la logique ou l'optimisation, soit elle implique de résoudre de nombreux problèmes distincts et complexes. En ce qui concerne les différents défis dans le domaine de l'IA, le problème central de création d'intelligence a été divisé en sous-problèmes qui concernent les capacités particulières dont un système intelligent devrait disposer. Les principaux sous-problèmes sont l'apprentissage automatique, la capacité des machines à planifier, raisonner, résoudre des problèmes, représenter des connaissances, la perception, la robotique, le traitement du langage naturel ou l'intelligence sociale²². Chacun de ces sous-problèmes correspond à un champ d'étude des sciences informatiques.

Enfin, dans l'histoire de l'IA, il est possible d'identifier un ensemble de méthodes au sens large, qui relèvent de trois catégories différentes :

1. La « cybernétique » étudie les liens entre neurobiologie et théorie de l'information et a pour but de développer des machines qui utilisent des réseaux électroniques pour afficher une intelligence rudimentaire²³ ;
2. L'IA « symbolique » se fonde sur l'hypothèse qu'il est possible de modéliser de nombreux aspects de l'intelligence humaine à partir de la manipulation de symboles²⁴ ;
3. L'IA « basée sur l'apprentissage statistique » s'appuie sur une approche avant tout mathématique.

Les méthodes bien connues sur lesquelles se fondent les systèmes d'IA sont : la « logique », pour la représentation des connaissances et la résolution de problèmes ; les « méthodes probabilistes », qui s'appliquent au raisonnement, à la planification, l'apprentissage, la perception, et la robotique ; les « méthodes de recherche et d'optimisation », utilisées pour la planification et la robotique ; les « méthodes d'apprentissage automatique », telles que la classification par arbres de décision, les machines à vecteurs de support (*support vector machines*) ; et les « réseaux de neurones profonds », qui sont utilisés pour presque tous les défis. L'inconvénient de certaines de ces puissantes méthodes d'apprentissage statistique est qu'elles ne peuvent être interprétées, c'est-à-dire qu'un humain ne peut comprendre la logique de la prise de décisions par de tels systèmes.

²² Poole, D., Mackworth, A., et Goebel, R., *op.cit.*

²³ Wiener, N., *Cybernetics (or control and communication in the animal and the machine)*, Cambridge (Massachusetts).

²⁴ Haugeland, J., *Artificial intelligence: the very idea*.

1.1.3. Applications de l'intelligence artificielle

Que peut faire l'IA aujourd'hui, et dans quels domaines est-elle utilisée ? Il n'est pas facile de répondre de façon exhaustive à cette question, dans la mesure où l'IA intervient aujourd'hui dans pléthore de tâches et de domaines. Dans la section suivante, nous faisons brièvement état de quelques applications d'IA qui pourraient présenter un intérêt particulier pour les lecteurs de cette publication.

- **Véhicules robotisés.** Les véhicules autonomes ou sans conducteur ont été rendus possibles grâce aux progrès de l'IA. Des composantes distinctes d'IA sont incorporées dans des systèmes tels que la prévention des collisions, le changement de file, le freinage, etc. qui contribuent au fonctionnement global des voitures autonomes. Les entreprises d'IA qui travaillent sur les véhicules robotisés sont Tesla, Google, et Apple²⁵.
- **Santé.** L'IA est utilisée dans le domaine de la santé pour soutenir les médecins. Par exemple, les systèmes d'IA peuvent être utilisés pour le diagnostic médical, l'analyse des relations entre traitements et résultats, l'identification de problèmes liés au dosage, ainsi que pour assister les chirurgiens pendant leurs opérations, aider les radiologues à interpréter des images et créer de nouveaux médicaments²⁶.
- **Marketing, économie et finances.** Les entreprises et les institutions financières ont été les premières à adopter des systèmes d'IA, pour analyser les marchés, prédire l'évolution de la clientèle, prévoir les prix, suivre les actions, gérer les portefeuilles, dans le cadre du trading algorithmique, etc. L'IA est également utile pour réduire la fraude et les délits financiers.
- **Médias.** L'analyse des contenus médiatiques tels que les programmes TV, les publicités, les films et les vidéos peut être rendue possible grâce à des applications d'IA spécifiques. Les usages habituels concernent la reconnaissance des visages et des objets, le sous-titrage automatique, la reconnaissance de scène et pour résumer les contenus. L'analyse de médias fondée sur l'IA permet de créer des mots-clés qui décrivent un contenu média de façon à simplifier les recherches. Une autre application concerne le contrôle du caractère approprié des contenus médiatiques ou la détection automatique de logos et produits appropriés/inappropriés dans un cadre publicitaire.
- **Information et édition.** De nombreuses entreprises utilisent désormais des techniques d'IA pour générer des informations et des articles automatiquement. Grâce à l'IA, les entreprises sont également en mesure de produire du texte. Un exemple d'application concerne la génération de résumés personnalisés d'événements sportifs²⁷. Une autre application convertit des données structurées en langage naturel.
- **Musique.** L'IA a permis, dans une certaine mesure, d'imiter la composition de musique par l'homme, et elle permet aussi d'aider les humains à jouer de la musique

²⁵ CBInsights, *33 Corporations working on autonomous vehicles*, consulté le 16 mars 2017.

²⁶ Coiera, E., *Guide to Medical Informatics, the Internet and Telemedicine*, Chapman & Hall, Ltd., GBR, 1^{ère} édition.

²⁷ Disponible sur : <https://www.barrons.com/articles/big-data-and-yahoos-quest-for-mass-personalization-1377938511>.



ou à chanter²⁸. Les technologies d'accompagnement par ordinateur sont capables d'écouter et de suivre un musicien de façon à jouer en synchronie. Les technologies de composition interactives permettent à l'IA de répondre avec une composition musicale à la performance en direct d'un musicien. Enfin, des projets tels que Google Magenta, Sony Flow Machines, ou IBM Watson Beat sont capables de composer de la musique de n'importe quel style après avoir analysé de larges bases de données de chansons. D'autres applications d'IA pour la musique couvrent également les activités de marketing et d'écoute de la musique.

- **Deepfakes.** Les *deepfakes* sont des contenus médias artificiels créés à partir de techniques d'IA et qui semblent réels pour le regard humain²⁹. Il s'agit généralement d'images ou de vidéos dans lesquelles une personne est remplacée par une autre grâce à des méthodes d'apprentissage profond. Les principales méthodes utilisées pour créer des *deepfakes* impliquent le développement de modèles génératifs tels que les réseaux antagonistes génératifs (*generative adversarial networks, GAN*)³⁰ ou les auto-encodeurs³¹. Bien que les *deepfakes* puissent être utilisés à des fins humoristiques, ils sont surtout connus pour produire des canulars, de fausses informations, des vidéos de célébrités à caractère pornographique et pour permettre des fraudes financières. Ainsi, tant les gouvernements que l'industrie développent des outils d'IA pour les détecter et les limiter. En effet, dans un futur proche, l'IA sera probablement capable non seulement de créer des images et des vidéos réalistes, mais aussi de véritables contenus médiatiques, par exemple des films, des séries TV et des programmes tels que des émissions de télé-réalité ou des quiz.

1.2. Qu'est-ce que l'intelligence artificielle explicable ?

De nos jours, les systèmes d'IA sont non seulement capables de simuler les processus de pensée et d'apprentissage humains, mais ils peuvent aussi dépasser l'intelligence humaine dans l'accomplissement de certaines tâches. Cela est possible parce que l'intelligence artificielle diffère de l'intelligence humaine et, en raison du recours généralisé à des méthodes complexes telles que l'apprentissage profond, l'IA ne se comporte pas comme l'intelligence humaine, ou du moins elle agit sur la base d'un processus décisionnel qui n'est pas toujours compréhensible par l'homme. En effet, la dernière décennie a vu l'avènement de ce que Frank Pasquale appelle la « société de la boîte noire³² », dans laquelle les systèmes d'IA s'appuient sur des modèles décisionnels obscurs pour prendre leurs décisions. Le choix résulte de la bonne performance de ces modèles à boîte noire en

²⁸ Roads, C., « Research in music and artificial intelligence », *ACM Computing Surveys (CSUR)*, 17(2):163–190.

²⁹ Kietzmann, J., Lee, L. W., McCarthy, I. P., et Kietzmann, T. C., « Deepfakes: Trick or treat? », *Business Horizons*, 63(2):135–146.

³⁰ Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., et Bengio, Y., « Generative adversarial nets », in *Advances in neural information processing systems*, pp. 2672–2680.

³¹ Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., et Frey, B., « Adversarial autoencoders », *arXiv preprint arXiv:1511.05644*.

³² Pasquale, F., *The black box society*, Harvard University Press.

termes de précision³³. Il s'agit par exemple de réseaux de neurones et de réseaux de neurones profonds, de machines à vecteurs de support, de combinaisons de classifieurs ensemblistes ou apprentissages d'ensemble (*ensemble classifiers*), mais aussi de compositions de systèmes expert, d'exploration de données et de logiciels codés en dur qui « cachent » la logique de leurs processus décisionnels internes aux humains³⁴. Ainsi, les modèles à boîte noire sont des outils sur lesquels s'appuie l'IA pour accomplir des tâches et dont le processus décisionnel n'est soit pas accessible, soit non compréhensible pour l'homme.

L'absence d'explication sur la manière dont ces modèles à boîte noire prennent leurs décisions pose un problème pour leur adoption dans des contextes critiques du point de vue de la sécurité ou qui sont sensibles au niveau social, tels que la santé ou la loi. Le problème ne se limite pas à l'absence de transparence mais aussi aux biais dont les boîtes noires pourraient hériter en raison de préjugés et d'artefacts qui seraient dissimulés au sein des données d'apprentissage utilisées par les modèles d'apprentissage automatique obscurs des systèmes d'IA. En effet, les algorithmes d'apprentissage automatique développent des modèles après une phase d'apprentissage qui est rendue possible par les données massives (ou big data) provenant des registres d'entreprises et des traces numériques que laissent les gens au quotidien (par exemple leurs achats, déplacements, messages sur les réseaux sociaux, etc.). Cette immense quantité de données peut contenir des biais et des préjugés humains. Ainsi, les modèles de prise de décision qui sont tirés de ces données peuvent hériter de ces biais, et peut-être entraîner des décisions injustes ou mauvaises. C'est pourquoi la recherche en IA explicable (IAE) a récemment reçu beaucoup d'attention³⁵.

De plus, le Règlement général sur la protection des données (RGPD)³⁶ introduit un droit pour tout un chacun d'obtenir « des informations utiles concernant la logique sous-jacente » dès lors qu'il s'agit de décisions automatisées. En dépit d'avis divergents de la part de spécialistes du droit quant à la portée réelle de cette clause³⁷, le fait qu'un tel principe est impératif et qu'il représente un immense défi scientifique encore non résolu fait l'objet d'un consensus.

³³ Tan, P.-N. *et al.*, *op.cit.*

³⁴ Les lecteurs qui le souhaitent peuvent consulter plus d'informations sur les réseaux de neurones, les machines à vecteurs de support et les classifieurs d'ensembles dans Tan, P.-N. *et al.*, *op.cit.*

³⁵ Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., et Pedreschi, D. (2018), *op.cit.* Miller, T., « Explanation in artificial intelligence : Insights from the social sciences », *Artificial Intelligence*, 267:1–38. Adadi, A. et Berrada, M., *Peeking inside the black-box: A survey on explainable artificial intelligence (xai)*. IEEE Access, 6:52138–52160.

³⁶ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la Directive 95/46/CE (Règlement général sur la protection des données). Disponible sur <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679&from=EN&lang3=choose&lang2=choose&lang1=FR>.

³⁷ Malgieri, G. et Comandé, G., « Why a right to legibility of automated decision-making exists in the General Data Protection Regulation », *International Data Privacy Law*, 7(4):243–265. Goodman, B. et Flaxman, S., « EU regulations on algorithmic decision-making and a 'right to explanation' », in *ICML workshop on human interpretability in machine learning (WHI 2016)*, New York, NY, http://arxiv.org/abs/1606.08813_v1. Wachter, S., Mittelstadt, B., et Floridi, L., « Why a right to explanation of automated decision-making does not exist in the general data protection regulation », *International Data Privacy Law*, 7(2):76–99.

L'IAE est au cœur d'une science responsable au sein de nombreux secteurs industriels et disciplines scientifiques. Comment les entreprises peuvent-elles faire confiance à leurs produits d'IA sans comprendre le raisonnement suivi par leurs composantes d'apprentissage automatique ? Et comment les utilisateurs pourraient alors faire confiance à des services d'IA ? Il est impossible d'améliorer la confiance des gens en l'IA sans expliquer la logique suivie par des modèles obscurs.

1.2.1. Motivations de l'IA

En plus des motivations théoriques, éthiques et juridiques qui justifient le besoin en IA explicable, il existe de véritables cas dans lesquels des erreurs ou une discrimination auraient pu être évitées si l'IA n'avait pas été obscure. L'accès aux motifs expliquant une décision par IA est particulièrement important en ce qui concerne les systèmes critiques du point de vue de la sécurité, par exemple pour les véhicules autonomes ou la médecine, dans lesquels une mauvaise décision pourrait aller jusqu'à provoquer la mort d'individus. Lorsqu'un Uber autonome a renversé et tué une passante à Tempe, Arizona en 2018, le recours à des modèles interprétables aurait aidé Uber à comprendre les raisons expliquant cette action, et à gérer sa responsabilité.

Un autre risque inhérent aux composantes de la boîte noire utilisées par les systèmes d'IA est la possibilité de prendre des décisions erronées, résultant de la présence de fausses corrélations ou d'artefacts dans les données d'apprentissage. Par exemple, Ribeiro *et al.*³⁸ ont montré qu'un classifieur entraîné à distinguer des loups de chiens husky fondait ses prédictions uniquement sur la présence de neige à l'arrière-plan. L'IA faisait ce choix parce que toutes les images figurant des loups utilisées pour l'apprentissage incluaient de la neige en arrière-plan. Dans un autre exemple, en 2016, le logiciel d'IA utilisé par Amazon pour décider des zones dans lesquelles Amazon offrait un service de livraison gratuite dans la journée aux États-Unis a involontairement exclu du programme les quartiers à dominante minoritaire (souvent alors même que tous les quartiers alentour étaient inclus)³⁹. Plus récemment, les journalistes de ProPublica ont démontré que le score COMPAS, un modèle prédictif du « risque de récidive criminelle » (propriété de Northpointe), avait un fort biais ethnique. En effet, selon ce score, une personne noire qui n'avait pas récidivé était deux fois plus souvent considérée comme « à haut risque » qu'une personne blanche qui n'avait pas récidivé. À l'inverse, les récidivistes blancs étaient classés, deux fois plus souvent que les récidivistes noirs, comme présentant un « faible risque »⁴⁰.

³⁸ Ribeiro, M. T., Singh, S., et Guestrin, C. (2016), « Why should I trust you ? : Explaining the predictions of any classifier », in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, ACM.

³⁹ Disponible sur : <http://www.techinsider.io/how-algorithms-can-be-racist-2016-4>.

⁴⁰ Disponible sur : <http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

1.2.2. Les dimensions de l'interprétabilité

« Interpréter » signifie donner un sens ou expliquer et présenter certains concepts en des termes compréhensibles⁴¹. Ainsi, l'« interprétabilité » de l'IA se définit comme la capacité à expliquer ses décisions ou à leur donner une signification, en des termes compréhensibles pour l'homme⁴². Cette définition implique que les concepts qui composent l'explication se suffisent à eux-mêmes et ne nécessitent pas d'explications supplémentaires. Schématiquement, une explication est une « interface » entre un humain et une IA, qui est compréhensible par l'homme et donne en même temps une représentation fidèle de l'IA. Une série de « dimensions » peuvent être identifiées pour analyser l'interprétabilité des systèmes d'IA, qui reflètent les différents types d'explications possibles⁴³.

1.2.2.1. Explication de la boîte noire ou explication dès la conception

Une distinction peut être établie entre l'explication de modèles à boîte noire et l'explication intégrée dès la conception d'un modèle d'IA. Dans le premier cas, le principe est de combiner un modèle d'IA à boîte noire avec une méthode d'explication capable d'interpréter les décisions de la boîte noire. Dans le second cas, la stratégie consiste à substituer au modèle obscur un modèle transparent dans lequel le processus décisionnel est conçu dès le départ pour être accessible. De façon plus précise, l'intérêt de l'explication du modèle à boîte noire est de permettre de conserver la bonne performance du modèle obscur d'IA tout en lui appliquant une technique d'IAE pour en extraire les explications⁴⁴. Ce type d'approche est le plus fréquent aujourd'hui dans le domaine de la recherche en IAE. De l'autre côté, l'« explication dès la conception » consiste à développer un modèle transparent qui est interprétable, et de substituer à la composante boîte noire du système d'IA un modèle interprétable⁴⁵. Les études sur le sujet font ressortir plusieurs modèles qui sont reconnus comme étant interprétables. Les exemples incluent les « arbres de décision », les « règles de décision » et les « modèles linéaires »⁴⁶. Ces modèles sont considérés comme faciles à comprendre et à interpréter pour les humains. Mais presque tous sacrifient la performance en faveur de l'interprétabilité. En outre, ils ne peuvent être efficacement appliqués à

⁴¹ Disponible sur : <https://www.merriam-webster.com/>.

⁴² Doshi-Velez, F. et Kim, B., « Towards a rigorous science of interpretable machine learning », *arXiv preprint arXiv:1702.08608*. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., *et al.*, « Explainable artificial intelligence (xai) : Concepts, taxonomies, opportunities and challenges toward responsible ai », *Information Fusion*, 58:82–115.

⁴³ Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., et Pedreschi, D. (2018), *op.cit.*

⁴⁴ Craven, M. et Shavlik, J. W., « Extracting tree-structured representations of trained networks », in *Advances in neural information processing systems*, pp. 24–30. Ribeiro, M. T., Singh, S., et Guestrin, C. (2016), *op.cit.* Lundberg, S. M. et Lee, S.-I., « A unified approach to interpreting model predictions », in *Advances in neural information processing systems*, pp. 4765–4774.

⁴⁵ Rudin, C., « Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead », *Nature Machine Intelligence*, 1(5):206–215. Rudin, C. et Radin, J., « Why are we using black box models in ai when we don't need to? a lesson from an explainable ai competition », *Harvard Data Science Review*, 1(2).

⁴⁶ Freitas, A. A., « Comprehensible classification models: a position paper », *ACM SIGKDD explorations newsletter*, 15(1):1–10.

certain types de données telles que les images ou le texte, mais seulement à des données tabulaires, relationnelles, en d'autres termes à des tableaux de données.

1.2.2.2. Explications globales ou locales

Cette distinction entre explication globale ou locale dépend de la capacité de l'explication à permettre la compréhension dans son ensemble de la logique sous-tendant un système d'IA, ou seulement d'un cas spécifique, c'est-à-dire l'interprétation d'une décision unique. Une explication « globale » donne le moyen d'interpréter toutes les décisions possibles d'un modèle à boîte noire. D'une façon schématique, le comportement de la boîte noire est imité par un modèle transparent exercé à se comporter de la même manière que le modèle à boîte noire tout en étant compréhensible par l'homme. En d'autres termes, le modèle interprétable qui imite la boîte permet d'obtenir une interprétation globale. Les explications globales sont assez difficiles à réaliser, et jusqu'à présent, elles ne peuvent être obtenues que dans le cadre d'une IA traitant des données tabulaires. Une explication locale consiste à extraire les motifs d'un « résultat » obtenu par un modèle à boîte noire dans le cas d'une décision unique. Dans ce cas, il n'est pas nécessaire d'expliquer toute la logique qui sous-tend l'IA, mais simplement la raison expliquant la prédiction obtenue dans le cas d'une entrée spécifique. Ainsi, un modèle interprétable est utilisé pour imiter le comportement de la boîte noire uniquement dans le « voisinage » du cas soumis à analyse, c'est-à-dire seulement dans des cas similaires. L'idée est qu'au sein de ce voisinage, il est plus facile d'imiter l'IA avec un modèle simple et interprétable. Plusieurs approches d'explications locales sont analysées dans les sections suivantes.

1.2.2.3. Modèles interprétables pour expliquer l'IA

Dans la section suivante, nous décrivons brièvement les modèles interprétables les plus souvent retenus pour expliquer les systèmes d'IA obscurs ou pour remplacer les composantes de la boîte noire.

- Un « arbre de décision » exploite un graphique, structuré comme un arbre et composé de nœuds internes qui constituent des tests de certains attributs et caractéristiques (par exemple, si une variable a une valeur inférieure, égale ou supérieure à un seuil donné) et de nœuds terminaux (les « feuilles » de l'arbre), qui représentent une décision. Chaque branche constitue un résultat possible⁴⁷. Les chemins allant des racines aux feuilles représentent les règles de classification. Les règles les plus communes sont de type « si-alors », dans lequel la proposition « si » est une combinaison de conditions concernant les variables d'entrée. Lorsque la proposition est vérifiée, la partie « alors » révèle l'action de l'IA.
- Dans le cas d'une « liste de règles », étant donné un ensemble ordonné de règles, l'IA renvoie comme décision la première règle qui est vérifiée⁴⁸.

⁴⁷ Quinlan, J. R., *C4.5: Programs for Machine Learning*, Elsevier.

⁴⁸ Yin, X. et Han, J., « Cpar: Classification based on predictive association rules », in *Proceedings of the 2003 SIAM International Conference on Data Mining*, pp. 331 – 335. SIAM.

- Enfin, les « modèles linéaires » permettent de visualiser l'importance de chaque attribut : tant le signe que la magnitude de la contribution des attributs pour une prédiction donnée⁴⁹. Si le signe de la valeur d'un attribut est positif, alors il contribue en augmentant le résultat obtenu, et il le réduit dans le cas contraire. Des magnitudes plus élevées des valeurs d'attribut indiquent une influence plus importante sur la prédiction du modèle.

1.2.2.4. Caractéristiques souhaitables de l'interprétabilité

Dans la mesure où les modèles interprétables sont nécessaires pour obtenir des explications, leur adoption devrait tenir compte de certaines caractéristiques souhaitables⁵⁰, afin d'accroître la confiance dans un modèle donné.

- L'« interprétabilité » consiste à évaluer dans quelle mesure une explication donnée est compréhensible pour l'homme. Une approche souvent retenue pour mesurer l'interprétabilité est la « complexité » du modèle de substitution interprétable. La complexité est habituellement estimée à partir de la « taille » du modèle interprétable. Par exemple, la complexité peut être mesurée au nombre de clauses dans la condition ; pour les modèles linéaires, il est possible de compter le nombre des « poids non nuls », tandis que pour les arborescences décisionnelles, il s'agit de la profondeur de l'arbre.
- La « fidélité » consiste à évaluer dans quelle mesure le modèle substitutif interprétable est capable d'« imiter » avec précision, de façon globale ou locale, la décision de l'IA. La fidélité peut être en pratique mesurée en termes de score de précision, de F-score, etc.⁵¹ des décisions prises par le modèle à boîte noire. De plus, un modèle substitutif interprétable devrait également répondre à d'autres conditions importantes générales : par exemple, être capable de prendre des décisions adaptées dans des situations inédites.
- Le « caractère équitable » et le « respect de la vie privée » sont indispensables pour garantir la protection de certains groupes contre les discriminations⁵² et pour s'assurer que les modèles interprétables ne divulguent pas d'informations sensibles⁵³.
- La « facilité d'utilisation » est une autre caractéristique qui peut influencer la confiance dans un modèle : par exemple, une explication interactive peut être plus utile qu'une explication textuelle prédéterminée.

⁴⁹ Ribeiro, M. T., Singh, S., et Guestrin, C. (2016), *op.cit.*

⁵⁰ Freitas, A. A., *op.cit.*

⁵¹ Tan, P.-N. *et al.*, *op.cit.*

⁵² Romei, A. et Ruggieri, S., « A multidisciplinary survey on discrimination analysis », *The Knowledge Engineering Review*, 29(5):582–638.

⁵³ Aldeen, Y. A. A. S., Salleh, M., et Razzaque, M. A., *A comprehensive review on privacy preserving data mining*. SpringerPlus, 4(1):694.

1.2.2.5. Méthodes explicatives spécifiques ou indépendantes du modèle

Une distinction peut être faite entre méthodes d'explication spécifiques ou indépendantes du modèle, selon que la technique retenue pour extraire l'explication agit sur un modèle particulier appliqué par un système d'IA, ou que celle-ci peut être appliquée à n'importe quel type d'IA. L'approche la plus utilisée pour expliquer les boîtes noires d'IA est connue sous le nom d'« rétro-ingénierie » (*reverse engineering*). Le terme indique que l'explication est tirée de l'observation du résultat, c'est-à-dire de la décision de l'IA, quand les entrées sont modifiées de façon contrôlée.

- Une méthode explicative est « spécifique à un modèle », ou non généralisable⁵⁴, si elle ne peut être utilisée que pour l'interprétation de certains types de modèles à boîte noire. Par exemple, si une approche explicative est conçue pour interpréter une forêt d'arbres de décision⁵⁵, et qu'elle s'appuie en interne sur la notion de distance entre les arbres, alors elle n'est pas applicable aux prédictions engendrées par un modèle de réseau de neurones.
- D'un autre côté, une méthode est « indépendante du modèle », ou généralisable, lorsqu'elle peut être utilisée indépendamment du modèle à boîte noire en question : les caractéristiques internes de l'IA ne sont pas exploitées pour développer le modèle interprétable chargé d'imiter le comportement de la boîte noire.

1.2.2.6. Niveau de connaissances de l'utilisateur

Des niveaux variables de connaissances de base et des expériences dans des tâches variées sont liés à des notions et des exigences différentes pour l'utilisation des explications. Les experts du domaine peuvent être en mesure de comprendre des explications complexes, tandis que des utilisateurs ordinaires requièrent des clarifications simples et efficaces. En effet, l'utilité et la pertinence d'une explication dépendent de l'utilisateur⁵⁶. Si l'on prend l'exemple du COMPAS mentionné plus haut, une explication spécifique du score peut avoir du sens pour un juge qui chercherait à comprendre et vérifier la suggestion du système d'IA et qui pourrait se rendre compte que celui-ci a un biais qui défavorise les personnes de couleur noire. À l'inverse, la même explication n'est pas utile à un prisonnier qui ne peut changer sa réalité de personne de couleur noire. En revanche, le prisonnier pourrait trouver utile, et donc porteuse de sens, la suggestion selon laquelle son risque de récidive sera jugé moindre lorsqu'il sera plus âgé, ce qui voudrait dire qu'une résidence surveillée pourrait alors lui être plus facilement accordée.

⁵⁴ Martens, D., Baesens, B., Van Gestel, T., et Vanthienen, J., « Comprehensible credit scoring models using rule extraction from support vector machines », *European journal of operational research*, 183(3):1466–1476.

⁵⁵ Tan, P.-N. *et al.*, *op.cit.*

⁵⁶ Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., Ghosh, J., Puri, R., Moura, J. M., et Eckersley, P., « Explainable machine learning in deployment », in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 648–657.

1.2.2.7. Limitations de temps

Le temps qu'un utilisateur est autorisé à consacrer à la compréhension d'une explication ou est disposé à le faire est également un aspect crucial. Évidemment, le temps dont l'utilisateur dispose est étroitement lié au scénario dans lequel le modèle prédictif doit être utilisé. Dans certains contextes, dans lesquels l'utilisateur doit rapidement prendre une décision, par exemple une opération chirurgicale ou en cas de désastre imminent, il est préférable d'avoir accès à une information simple et efficace. À l'inverse, dans les contextes où la contrainte de temps est moindre, par exemple au cours d'une procédure de demande de prêt, l'utilisateur pourrait souhaiter bénéficier d'une explication plus complexe et exhaustive.

1.2.3. Comment lire les différentes explications

Le domaine émergent de l'IAE fait naître un large éventail de possibilités pour expliquer les composantes de boîte noire des systèmes d'IA. En effet, il n'est pas possible de concevoir un type d'explication unique, qui serait applicable à toutes les applications. Les sections suivantes illustrent les types d'explications les plus utilisés.

1.2.3.1. Explications globales

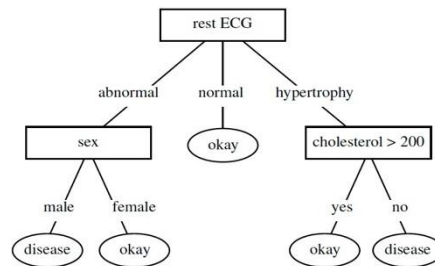
1.2.3.1.1. Explications par arborescence décisionnelle

L'imitation d'une composante obscure d'IA par une arborescence décisionnelle a été l'une des premières approches mises en œuvre⁵⁷. La méthode TREPAN est capable de représenter au moyen d'une seule arborescence décisionnelle toutes les décisions possibles que peut prendre un réseau de neurones à partir de données tabulaires. TREPAN construit une arborescence décisionnelle qui imite les concepts représentés par les réseaux en maximisant le rapport de gain⁵⁸ calculé à partir de la fidélité de l'arborescence pour la décision du réseau de neurones. Les résultats de TREPAN permettent l'exploration globale d'un réseau de neurones à partir d'une structure d'arborescence qui, depuis la racine et pour chaque voie possible, indique les conditions expliquant la décision retenue par le système d'IA.

⁵⁷ Craven, M. et Shavlik, J. W., *op.cit.*

⁵⁸ Tan, P.-N *et al.*, *op.cit.*

Illustration 1. Exemple d'explication globale par arborescence décisionnelle obtenue avec TREPAN



1.2.3.1.2. Liste de règles

Comme nous l'avons mentionné plus haut, une autre piste pour expliquer la classification par boîte noire est de concevoir dès le départ des modèles transparents pour les systèmes d'IA. La méthode CORELS⁵⁹ fournit une liste de règles, et constitue une solution optimale pour les données tabulaires. Un exemple de liste de règles figure au schéma 2. Les règles sont lues les unes après les autres, et l'IA prend la décision correspondant à la première règle pour laquelle les conditions sont vérifiées.

Illustration 2. Exemple d'explication sous forme de liste de règles obtenue avec CORELS

```

if (age=23-25) and (priors=2-3) then predict yes
else if (age = 18 - 20) then predict yes
else if (sex = male) and (age = 21 - 22) then predict yes
else if (priors > 3) then predict yes
else predict no
  
```

1.2.3.2. Explications locales

Les explications ci-dessus sont globales. Toutefois, lorsque les modèles d'IA obscurs sont trop complexes, il est préférable d'adopter une méthode d'IAE locale et d'extraire séparément les raisons expliquant les décisions pour chaque cas donné. C'est pourquoi la recherche en IAE se concentre davantage de nos jours sur les explications locales. Les explications locales les plus représentatives sont décrites dans les sections suivantes.

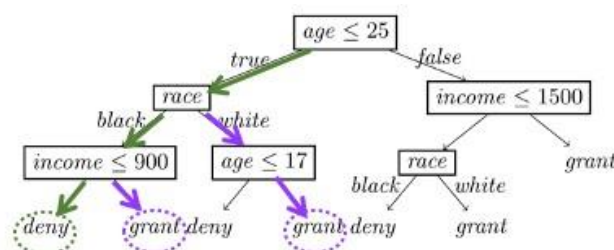
1.2.3.2.1. Explications basées sur des règles

Dans le cas des explications de type « si-alors » (« si telles conditions sont remplies, alors on obtient tel résultat »), le « résultat » correspond à la décision de l'IA, tandis que les « conditions » expliquent les « raisons factuelles » du résultat. Par exemple, l'explication

⁵⁹ Angelino, E., Larus-Stone, N., Alabi, D., Seltzer, M., et Rudin, C., « Learning certifiably optimal rule lists », in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 35–44. ACM.

du refus d'une demande de prêt d'un client correspondant à « âge=22, race=noir et revenu=800 » par une banque qui utilise l'IA pourrait représenter la règle factuelle « si $\text{âge} \leq 25$ et $\text{race} = \text{noir}$ et $\text{revenu} \leq 900$ alors refuser ». La méthode LORE construit une arborescence décisionnelle locale dans le voisinage du cas analysé⁶⁰, puis en extrait une règle unique qui révèle les motifs d'une décision dans le cas particulier en question (voir le chemin vert sur le schéma 3). ANCHOR⁶¹ est un autre exemple d'approche d'IAE pour expliquer localement l'IA à partir de règles de décision appelées des ancres. Une ancre contient une série de caractéristiques avec les valeurs qui sont fondamentales pour l'obtention d'une certaine décision.

Illustration 3. Exemple d'explication à base de règles factuelles et contrefactuelles obtenue avec LORE



1.2.3.2.2. Importance des caractéristiques

Les explications locales peuvent aussi être utilisées pour obtenir l'importance des caractéristiques. Le schéma 4 montre l'importance des caractéristiques obtenue par LIME⁶², qui met en lumière les contributions positives et négatives prises en compte dans les résultats de la boîte noire et leur assigne une importance. LIME utilise un modèle linéaire comme substitut local interprétable et permet d'obtenir l'importance des caractéristiques comme explication en exploitant les coefficients de régression. Le schéma 5 montre l'importance des caractéristiques issues de SHAP⁶³. SHAP permet d'obtenir l'importance d'une caractéristique supplémentaire unique locale pour un cas spécifique. Plus une valeur SHAP est élevée, plus la contribution de cette caractéristique est importante. Utilisés dans le cadre de paramètres adaptés, LIME et SHAP peuvent aussi permettre d'expliquer les systèmes d'IA travaillant sur du texte.

⁶⁰ Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., et Turini, F. (2019a), « Factual and counterfactual explanations for black box decision making », *IEEE Intelligent Systems*.

⁶¹ Ribeiro, M. T., Singh, S., et Guestrin, C. (2018), « Anchors: High-precision model-agnostic explanations », in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*.

⁶² Ribeiro, M. T., Singh, S., et Guestrin, C. (2016), *op.cit.*

⁶³ Lundberg, S. M. et Lee, S.-I., *op.cit.*

Illustration 4. Exemple d'explication basée sur l'importance des caractéristiques obtenue avec LIME

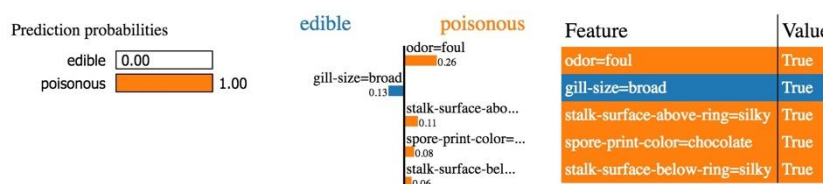


Illustration 5. Exemple d'explication basée sur l'importance des caractéristiques obtenue avec SHAP



1.2.3.2.3. Cartes de saillance

En traitement de l'image, les explications consistent habituellement en des « cartes de saillance », c'est-à-dire des images qui montrent la contribution positive (ou négative) de chaque pixel au résultat obtenu par la boîte noire. Les cartes de saillance sont conçues intelligemment pour expliquer localement les modèles de RNP par des méthodes d'attributions fondées sur le gradient et la perturbation. Ces approches d'IAE recherchent les pixels les plus importants d'une image de façon à maximiser la probabilité que l'IA donne la même réponse sans prendre en considération les pixels non pertinents. Dans le cadre de transformations d'images exploitant le concept de « super-pixels », des méthodes telles que LORE et LIME peuvent également être utilisées pour expliquer l'IA appliquée aux images. La méthode ABELE⁶⁴ utilise des modèles génératifs pour obtenir une carte de saillance qui met en lumière les aires contiguës qui peuvent être variées tout en maintenant une même décision de la part de la boîte noire. Le schéma 6 propose une comparaison des cartes de saillance pour la classification des chiffres 9 et 0 manuscrits, dans le cadre des méthodes d'explication ABELE,⁶⁵ LIME⁶⁶, SALiency⁶⁷, GRADInput⁶⁸, INTGrad⁶⁹ et ELRP⁷⁰.

⁶⁴ Guidotti, R., Monreale, A., Matwin, S., et Pedreschi, D. (2019b), « Black box explanation by learning image exemplars in the latent feature space », in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 189–205. Springer.

⁶⁵ Guidotti, R., Monreale, A., Matwin, S., et Pedreschi, D. (2019b), *op.cit.*

⁶⁶ Ribeiro, M. T., Singh, S., et Guestrin, C. (2016), *op.cit.*

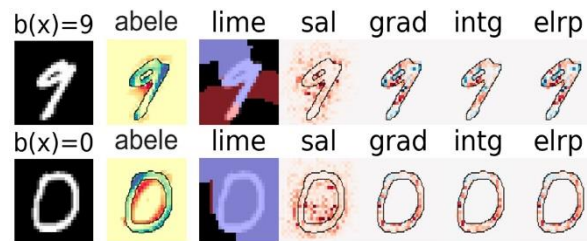
⁶⁷ Simonyan, K., Vedaldi, A., et Zisserman, A., « Deep inside convolutional networks: Visualising image classification models and saliency maps », *arXiv preprint arXiv:1312.6034*.

⁶⁸ Shrikumar, A. *et al.*, « Not just a black box: Learning important features through propagating activation differences », *arXiv:1605.01713*.

⁶⁹ Sundararajan, M. *et al.*, « Axiomatic attribution for dnn », in *ICML. JMLR. Tan, P.-N. et al.*

⁷⁰ Bach, S., Binder, A., *et al.*, « On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation », *PLoS one*, 10(7):e0130140.

Illustration 6. Exemple de cartes de saillance obtenues par différentes méthodes explicatives. La première colonne montre l'image analysée et le label assigné par le modèle à boîte noire b du système d'IA.



1.2.3.2.4. Explications basées sur des prototypes

Une explication basée sur des « prototypes » permet d'obtenir des spécimens similaires aux cas analysés, qui explicitent les motifs de la décision du système d'IA. Les prototypes sont utilisés comme le fondement de la représentation d'une catégorie ou d'un concept⁷¹. Les explications basées sur des prototypes peuvent porter sur des données tabulaires, des images ou du texte. Selon Li *et al.*⁷² et Chen *et al.*⁷³, les prototypes d'image sont utilisés comme fondement pour le concept d'interprétabilité⁷⁴. Kim *et al.*⁷⁵ s'interrogent sur le concept de « contre prototype » pour les données tabulaires, c'est-à-dire des prototypes qui montrent ce qui devrait être modifié pour obtenir une décision différente. Des exemples et des contre-exemples sont utilisés par ABELE⁷⁶ pour améliorer l'utilité de l'explication tirée d'une carte de saillance. Des exemples (à gauche) et des contre-exemples (à droite) pour les chiffres 9 et 0 figurent au schéma 7.

⁷¹ Frixione, M. et Lieto, A., « Prototypes vs exemplars in concept representation », in *KEOD*, pp. 226–232.

⁷² Li, O., Liu, H., Chen, C., et Rudin, C., « Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions », in *Thirty-second AAAI conference on artificial intelligence*.

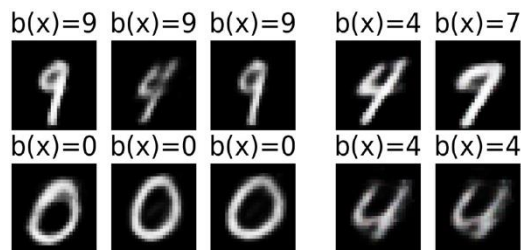
⁷³ Chen, C., Li, O., Barnett, A., Su, J., et Rudin, C., « This looks like that: deep learning for interpretable image recognition », *arXiv:1806.10574*.

⁷⁴ Bien, J. et Tibshirani, R., « Prototype selection for interpretable classification », *The Annals of Applied Statistics*, 5(4):2403–2424.

⁷⁵ Kim, B., Koyejo, O. O., et Khanna, R., « Examples are not enough, learn to criticize! criticism for interpretability », in *Advances In Neural Information Processing Systems*, pp. 2280–2288.

⁷⁶ Guidotti, R., Monreale, A., Matwin, S., et Pedreschi, D. (2019b), *op.cit.*

Illustration 7. Exemple d'explication par exemples (gauche) et contre-exemples (droite) obtenue avec ABELE. Au-dessus de chaque (contre-)exemple figure le label attribué par le modèle à boîte noire b du système d'IA.



1.2.3.2.5. Explications contrefactuelles

Une explication « contrefactuelle » montre ce qui devrait être différent pour modifier la décision du modèle à boîte noire. Les explications contrefactuelles sont importantes en ce sens qu'elles aident les personnes à raisonner sur les relations de cause à effet entre les caractéristiques observées et les résultats de classification obtenus⁷⁷. Bien que les explications directes et factuelles, telles que les règles de décision ou l'importance des contributions, soient cruciales pour comprendre les motifs d'un certain résultat, une explication contrefactuelle révèle ce qui devrait changer dans une instance donnée pour obtenir un résultat de classification différent⁷⁸. La méthode LORE⁷⁹ mentionnée ci-dessus fournit, en plus d'une règle d'explication factuelle, une série de *règles contrefactuelles*. Sur le schéma 3, la série de règles contrefactuelles figure en violet, et indique « si revenu $\geq$ 900 alors accord, ou si race = blanc alors accord », précisant quels changements renverseraient la décision. La méthode d'explication ABELE⁸⁰ propose des contre-exemples d'images qui révèlent les similitudes et les différences entre des instances de même classe et d'autres classes.

1.3. IA et IA explicable dans le domaine des médias

Les technologies d'IA transforment et réinventent l'industrie des médias ainsi que son modèle commercial, notamment en facilitant la monétisation des contenus et en fournissant aux utilisateurs finaux des services et publicités hyperpersonnalisés. Le recours à des applications d'IA est fortement répandu dans le cinéma, la télévision, la radio, la

⁷⁷ Byrne, R. M., « Counterfactuals in explainable artificial intelligence (xai): evidence from human reasoning », in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 6276–6282. Apicella, A., Isgrò, F., Prevete, R., et Tamburrini, G., « Contrastive explanations to classification systems using sparse dictionaries » in *International Conference on Image Analysis and Processing*, pp. 207–218. Springer.

⁷⁸ Wachter, S., Mittelstadt, B., et Floridi, L., *op.cit.*

⁷⁹ Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., et Turini, F. (2019a), *op.cit.*

⁸⁰ Guidotti, R., Monreale, A., Matwin, S., et Pedreschi, D. (2019b), *op.cit.*

presse écrite et la publicité. Selon Chan-Olmsted⁸¹, on peut distinguer deux caractéristiques pour les applications d'IA au sein du champ des médias :

- Certaines applications s'adressent plutôt au public des médias, le côté de « la demande », et d'autres se préoccupent davantage de la stratégie des fournisseurs de médias, soit de « l'offre ». Certaines concernent néanmoins ces deux groupes, par exemple, dans le domaine de l'interaction avec le public, l'expérience augmentée, et l'optimisation des messages.
- Certaines applications d'IA s'adressent aux « créateurs de contenus », alors que d'autres concernant l'analyse et la découverte des contenus visent plutôt les « distributeurs de contenus ». Les entreprises de loin les plus en avance à ce jour en termes d'adoption des technologies d'IA sont les services d'informations en ligne d'organisations telles que le *New York Times* et les services de vidéo à la demande tels que Netflix ou Prime Video. Ces applications d'AI concernent par exemple les recommandations, la personnalisation, l'écoute et le suivi des réseaux sociaux, le suivi de l'état émotionnel, l'accessibilité, la création et la postproduction de vidéos, la vérification d'information, l'analyse prédictive, les relations clients, la rédaction automatisée et les assistants vocaux.

1.3.1. Applications d'IA et explicabilité

Les entreprises technologiques telles qu'Amazon, Netflix, Facebook, et Google sont les principaux moteurs de l'expansion de l'IA dans le secteur des médias. Par exemple, les « systèmes de recommandations » de Amazon Prime, Netflix et Spotify se fondent sur des méthodes d'IA⁸². Une étude récente a montré que l'IA était avant tout exploitée par les nouveaux médias pour améliorer les services de recommandations. Une autre application de l'IA pourrait être de réinventer la connexion média-public, c'est-à-dire que l'IA pourrait être utilisée pour comprendre les sentiments et préférences du public ainsi que ses conversations sociales, ce qui rendrait possible de s'adapter en temps réel à l'intérêt du public et de permettre une expérience de consommation améliorée grâce à des contenus médiatiques personnalisés. Enfin, l'IA pourrait aider les entreprises des médias à identifier de nouvelles opportunités commerciales : des intrigues ou des personnages pourraient être créés à partir des préférences et des goûts des utilisateurs, de leurs opinions sur les médias sociaux, de leurs conversations, etc.⁸³. Dans la section suivante, nous passons en revue certaines applications spécifiques d'IA dans le domaine des médias.

⁸¹ Chan-Olmsted, S. M., « A review of artificial intelligence adoptions in the media industry », *International Journal on Media Management*, 21(3-4):193– 215.

⁸² Chan-Olmsted, S. M., *op.cit.*

⁸³ Kietzmann, J., Paschen, J., et Treen, E., « Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey », *Journal of Advertising Research*, 58(3):263–267.

1.3.1.1. Recommandations

L'usage le plus répandu de l'IA dans le champ médiatique concerne la recommandation de contenus. L'objectif d'un « système de recommandations » est de prédire le « rang » ou la « préférence » qu'un utilisateur serait susceptible d'accorder à un certain contenu par rapport aux autres. Les systèmes de recommandations reposent généralement sur un « filtrage collaboratif » et un « filtrage basé sur le contenu », ainsi que sur d'autres systèmes tels que des systèmes fondés sur la connaissance⁸⁴. Les systèmes de recommandations ont été largement adoptés dans de nombreux domaines, mais le champ des médias est celui qui correspond le mieux à leur usage. La logique de ces approches consiste à modéliser le comportement passé d'un utilisateur vis-à-vis des contenus médiatiques qu'il a précédemment « sélectionnés » ou auxquels il a donné des notes chiffrées. De plus, les systèmes de recommandations peuvent également prendre en compte les comportements similaires affichés par d'autres utilisateurs.

L'usage fortement répandu des systèmes de recommandations et le besoin de développer la confiance des utilisateurs dans les systèmes d'IA signifient que chaque utilisateur doit avoir accès à des recommandations explicables⁸⁵. En d'autres termes, les recommandations ne doivent pas se contenter d'être exactes et utiles, elles doivent également être compréhensibles. Les types les plus pertinents de recommandations explicables dans le domaine des médias sont les explications fondées sur les utilisateurs, sur les caractéristiques, ou sur les éléments :

- Les explications fondées sur les utilisateurs peuvent prendre la forme suivante : « ce contenu vous est recommandé parce que des utilisateurs similaires l'ont sélectionné par le passé ». Elles se composent d'une série d'utilisateurs similaires (anonymisés) accompagnée des contenus qu'ils ont sélectionnés.
- Les explications fondées sur les caractéristiques seraient du type : « ce contenu vous est recommandé parce que sa description comprend des caractéristiques (par exemple les sujets, acteurs, musique, etc.) que vous aimez », et les propose selon le rang que vous leur avez assigné.
- Enfin, les explications fondées sur les éléments indiqueraient : « ce contenu vous est recommandé parce qu'il est similaire à d'autres contenus que vous avez aimés par le passé⁸⁶ ».

Les anciens modèles de recommandations, tels que les modèles fondés sur les utilisateurs ou les éléments, sont transparents et explicables. Une meilleure transparence est fondamentale pour améliorer la fiabilité, l'efficacité, la capacité de persuasion ainsi que la satisfaction de l'utilisateur final⁸⁷. Les avancées récentes en IA et le recours aux RNP ont amélioré la précision des recommandations, mais ont complètement éliminé toute

⁸⁴ Manning, C. D., Raghavan, P., et Schütze, H., *Introduction to information retrieval*, Cambridge university press.

⁸⁵ Zhang, Y. et Chen, X., « Explainable recommendation: A survey and new perspectives », *arXiv preprint arXiv:1804.11192*.

⁸⁶ Les lecteurs qui le souhaitent peuvent prendre connaissance d'autres types de recommandations explicables dans Zhang, Y. et Chen, X., *op.cit.*

⁸⁷ Tintarev, N. et Masthoff, J., « A survey of explanations in recommender systems », in *2007 IEEE 23rd international conference on data engineering workshop*, pp. 801–810. IEEE.

transparence, en raison de l'utilisation de modèles complexes obscurs tels que les RNP. Le manque d'explicabilité des systèmes de recommandations dans l'industrie des médias peut soulever de nombreux problèmes. Si le système n'indique pas aux utilisateurs pourquoi de tels résultats sont fournis, il pourrait être moins efficace pour les orienter vers un contenu particulier, ce qui pourrait porter davantage préjudice à la fiabilité du système.

1.3.1.2. Personnalisation et individualisation

La personnalisation de la sélection et de la mise à disposition de contenus médiatiques est un objectif fondamental pour lequel l'industrie des médias s'appuie sur l'IA. En effet, les systèmes d'IA dans les médias sont particulièrement efficaces pour moduler précisément la stratégie de distribution des contenus grâce aux systèmes de recommandations. Par exemple, les systèmes d'IA peuvent analyser les tendances sur les médias sociaux pour identifier les meilleurs contenus à diffuser. Une autre application consiste à analyser le public pour générer automatiquement des titres/résumés/illustrations accompagnés de mots-clés qui assurent une meilleure visibilité au contenu. De plus, l'IA peut automatiser la production de contenus médiatiques et leur sélection, mettre régulièrement à jour des listes de lecture thématiques, et profiler les utilisateurs pour proposer des recommandations individualisées. De cette manière, le contenu médiatique proposé à chaque utilisateur peut être différent et adapté au profil particulier de celui-ci, à son trajet quotidien, ainsi qu'à l'heure et au lieu où le service de média est utilisé. D'autres applications visent à impliquer l'utilisateur avec les bons contenus, proposés dans le bon format, au moment opportun et d'une façon complètement personnalisée. Il s'agit en quelque sorte d'un éditeur personnalisé qui offrirait à chaque individu la meilleure expérience possible.

1.3.1.3. La création de contenu

Comme nous l'avons mentionné précédemment, l'une des applications les plus récentes de l'IA concerne la création d'informations, de musique et de vidéos. Dans le cadre de ce qu'on appelle le « robot journalisme » ou le « journalisme automatisé », des systèmes d'IA utilisent des algorithmes de génération automatique de langage naturel pour transformer des données et connaissances en articles, images et vidéos d'actualité. Par exemple, les systèmes d'IA peuvent aisément rédiger des articles quelque peu rébarbatifs pour les humains, par exemple des bulletins météo ou financiers, sur la base d'articles précédents et des données disponibles. S'agissant de la vidéo, en exploitant la reconnaissance des images, l'IA peut générer des montages vidéo cohérents. La plupart des grands fournisseurs de logiciel d'édition proposent déjà des fonctions de traitement automatisé des vidéos pour faire gagner du temps aux éditeurs. D'autres programmes logiciels tels que Gingalab⁸⁸ se servent de l'IA pour créer des vidéos de type « best of », sur la base de lignes éditoriales prédéfinies (par exemple humour, tension, ou portant sur un protagoniste particulier, etc.). En septembre 2018, la BBC a diffusé un programme entièrement créé par un robot⁸⁹.

⁸⁸ Disponible sur : <https://gingalab.com/>.

⁸⁹ Disponible sur : <https://www.bbc.co.uk/programmes/b0bhwk3p#:~:text=Made%20by%20Machine%3A%20When%20AI%20Met%20the%20Archive.Documentary>.

1.3.1.4. La détection de faux contenus

Le principal point faible du succès incroyable de l'IA est d'avoir entraîné le phénomène des *deepfakes*⁹⁰. Heureusement, bien que l'IA puisse en effet générer de faux contenus médiatiques, elle peut aussi contribuer à les détecter. Ainsi, l'IA peut être un atout crucial dans la lutte contre la désinformation puisque la technologie qui fabrique des faux peut également être exploitée pour les identifier. En s'appuyant sur des capacités d'analyse étendues et des algorithmes d'apprentissage automatique, l'IA peut en partie automatiser la vérification de contenus médias tels que des informations, des images et des vidéos. Le principal problème vient du fait que la qualité de cette détection découle de l'expérience de l'IA, qui elle-même dépend de la disponibilité de sources de données permettant de distinguer entre vrais et faux contenus. En plus, une telle source d'information doit être générée par des humains, qui vont manuellement annoter les contenus médias en les étiquetant comme vrais ou faux. Cette étape manuelle peut créer un biais dans les données car les humains pourraient ne pas être en mesure de vérifier tout le contenu médiatique nécessaire à l'apprentissage d'un système d'IA entièrement fonctionnel. Les humains peuvent en outre être amenés à se fier à leur ressenti pour déterminer ce qui est réel ou faux. L'IAE peut être cruciale lors de cette étape d'apprentissage pour deux motifs. D'abord, ceux qui utilisent l'IA pour détecter les faux contenus veulent être certains que la logique suivie par l'IA pour identifier ces contenus est compréhensible pour l'homme. Ils attendent par exemple une explication du type : « cette information est fausse parce que les phrases sont trop répétitives et que les images affichées viennent d'autres sites internet ». Deuxièmement, il ne faut pas que les systèmes d'IA se fient à des ensembles de données biaisées pour fournir des suggestions. Si toutes les informations réelles viennent de la même source, l'explication pourrait être du type : « cette information est fausse parce qu'elle n'est pas partagée par le *New York Times* ».

1.3.1.5. Autres applications

L'IA et l'IAE peuvent être utilisées par l'industrie des médias pour de nombreuses autres applications. Dans la section suivante, nous en mentionnons certaines sans entrer dans le détail. L'IA peut être utilisée comme outil pour améliorer les conversations sur internet, en d'autres termes pour reconnaître le discours de haine, la discrimination, les trolls, etc. Dans ce cas également, il est nécessaire d'avoir accès aux raisons motivant l'identification de messages comme étant inappropriés. L'IA pour la reconnaissance vocale constitue la base de nombreux services modernes et des assistants vocaux tels qu'Alexa d'Amazon, Google Home, ou Siri de Apple, qui sont présents dans tous les équipements intelligents. Ils exploitent l'IA et le traitement du langage naturel pour répondre à nos questions et accomplir des tâches. Enfin, il est utile de mentionner que l'IA a des conséquences stratégiques pour la monétisation des contenus médiatiques et pour prédire leur succès. Dans le même temps, s'agissant de la déontologie des médias, l'IAE devient cruciale pour expliquer au public, d'une manière transparente, la logique que suivent les systèmes d'IA qui interagissent avec les utilisateurs et prennent des décisions pour eux. Certaines

⁹⁰ Voir section 1.1.3 ci-dessus pour une définition des deepfakes.

questions pourraient être soulevées, et grâce à l'IAE, les utilisateurs pourraient obtenir des réponses : par exemple, quelle est la bonne proportion entre personnalisation et découverte de contenus ? Quel niveau de recommandation est souhaitable ? Pourquoi ce contenu médiatique est-il considéré comme réel ? En vertu du RGPD, la première étape, pour l'industrie des médias, est de révéler clairement quels contenus sont recommandés/créés par une IA.

1.3.2. Les services de vidéo à la demande en pratique

Les services de vidéo à la demande ont transformé la façon dont nous consommons les contenus médiatiques, des séries TV aux films en passant par les programmes de comédie et les dessins animés. Au moyen d'algorithmes, ces services adaptent l'expérience de l'utilisateur grâce à une personnalisation poussée qui se fonde sur un large ensemble de métadonnées (notamment le genre, la catégorie, le casting et la date de sortie), les données comportementales de l'utilisateur (par exemple les recherches, la navigation, le score attribué, le type de terminal), mais aussi sur les rangées sélectionnées sur la page d'accueil, les titres sélectionnés pour ces rangées, les visuels associés à chaque film, les films de la liste de lecture, etc. Le système d'IA sous-jacent évolue en recueillant constamment les données personnelles de chaque utilisateur et en offrant toujours une visualisation individualisée des options sur lesquelles celui-ci est le plus susceptible de cliquer, en fonction de leur usage et du contexte. L'objectif final est de trouver la meilleure combinaison possible de contenus à même de satisfaire les utilisateurs, plutôt que simplement des contenus correspondant au plus grand nombre d'utilisateurs. Les algorithmes soutiennent donc la créativité et la diversité plutôt que la standardisation. Ces possibilités remarquables sont offertes sur la base de systèmes de recommandations d'IA recueillant les données de millions d'utilisateurs qui regardent et notent les contenus de ces plateformes.

Personne ne sait exactement comment fonctionnent ces systèmes de recommandations. Les services de vidéo à la demande donnent habituellement une description générale de leur système de recommandations en langage simple, mais ils ne révèlent pas les détails de ce processus. En ce sens, ces systèmes de recommandations sont des modèles à boîte noire par excellence. En théorie, un utilisateur pourrait ne pas s'intéresser à la façon dont les recommandations sont faites, dans la mesure où il souhaite juste se détendre devant un contenu médiatique divertissant. Cependant, ces recommandations pourraient ne pas être entièrement personnelles, mais être motivées par des stratégies de marketing ou, pire encore, être influencées par des biais dans les données utilisées pour le développement des modèles d'apprentissage automatique. Lorsqu'elles s'appuient sur une IA basée sur des systèmes de recommandations obscurs découlant de modèles d'apprentissage automatique, les décisions ou les prédictions reposent sur les apprentissages tirés de processus automatisés, et le jeu de données d'apprentissage disponible ou sélectionné peut ne pas être représentatif de la population qu'il est censé évaluer. Par exemple, des statistiques fondées sur des gens qui évitent des films avec des personnages principaux asiatiques pourraient résulter en une discrimination contre cette catégorie de films et leur donner à tort un score faible pour une population qui pourrait

néanmoins être intéressée par ce type de contenu médiatique. Ainsi, le recours à l'IAE pour comprendre les données d'apprentissage et analyser comment elles affectent les résultats pour des populations différentes est fondamental pour identifier les biais. Pour chacun de ces services, une explication globale pourrait décrire comment l'algorithme se comporte d'une manière générale. On pourrait par exemple être amené à découvrir que le système de recommandations de l'IA ne suggérera pas un film de trois heures juste avant minuit un soir de semaine. À l'inverse, une explication locale des services de vidéo à la demande pourrait décrire comment l'IA se comporte pour un individu donné. Par exemple, si le client en question regarde habituellement un contenu de vidéo à la demande de 12h00 à 14h00 pendant sa pause déjeuner, au travail, alors le service en question ne suggérera pas un film de trois heures à ce moment-là, parce qu'il pourrait avoir déduit que c'est le meilleur choix possible sur la base de la routine, alors même que ce film de trois heures correspondrait entièrement aux intérêts de l'utilisateur. D'autre part, peut-être que l'utilisateur ne souhaite pas que le service en question exploite ce type d'informations personnelles pour faire des recommandations. Ainsi, des explications détaillées pourraient aider les services de vidéo à la demande à renforcer la confiance de leurs utilisateurs. En théorie, pour toute application qui utilise l'IA dans le domaine des médias, les suggestions devraient être fondées sur des recommandations non biaisées et une relation de confiance entre le service et ses utilisateurs.

1.4. Conclusion

L'intelligence artificielle ne peut être la solution définitive de toute application, et en particulier en ce qui concerne le domaine des médias, elle doit être rattachée à un être humain, pour créer et vérifier les contenus médiatiques, mais aussi pour regarder les contenus recommandés. En effet, l'IA est fondamentale s'agissant de la demande et de l'accès aux contenus, ainsi que de leur monétisation. L'IA a un grand potentiel de bien social, dans la mesure où elle aide à se repérer dans des masses de contenus, optimise les recherches et les recommandations personnalisées et prévient les manipulations. Avec des outils appropriés d'IAE et un niveau de confiance adapté du public et des fournisseurs, l'IA pourrait en pratique donner un véritable

Les données massives

*L'obtention et l'exploitation des données personnelles par des parties tierces, qu'elles aient été fournies délibérément ou par inadvertance par les utilisateurs, peuvent également avoir un effet très intrusif dans leur vie personnelle. De plus, dans certaines situations, l'État ou des parties privées pourraient souhaiter obtenir des informations sur la vie d'un utilisateur qui vont plus loin que celles que celui-ci serait prêt à céder. Dans sa contribution à cette publication, **Andrea Pin** affirme que « le vaste déploiement de l'IA exige que la sphère médiatique prenne conscience de son rôle unique et que le secteur des médias devrait s'efforcer d'utiliser l'IA de manière légale, éthique et robuste. » La question du rôle que les plateformes de médias devraient jouer dans la gestion de leurs contenus fait l'objet d'une attention particulière. Les débats consistant à établir dans quelle mesure celles-ci devraient ne pas se limiter à un rôle purement passif afin de poursuivre « l'objectif louable de surveiller leur contenu », se poursuivent. À cet égard, le « manque d'humanité » de l'IA est précisément l'un de ses principaux défauts. Les algorithmes de filtrage sont extrêmement efficaces pour trouver et retirer des contenus potentiellement préjudiciables, mais ils ne peuvent se mesurer aux humains lorsqu'il s'agit de prendre des décisions nuancées dans des domaines juridiques complexes⁹¹.*

⁹¹ Barker A., Murphy H., « YouTube reverts to human moderators in fight against misinformation », *Financial Times*, 20 septembre 2020, <https://www.ft.com/content/e54737c5-8488-4e66-b087-d1ad426ac9fa>.

2. Les données massives ou la matière dont sont faits les rêves de l'IA

Andrea Pin, professeure associée de droit public comparé, Université de Padoue

2.1. Introduction

Il est communément admis que les données massives sont le pétrole de la révolution de l'IA⁹². Depuis que la science des données et l'ingénierie technologique ont uni leurs forces, un flux massif d'informations a inondé le globe, affectant notre façon de vivre et de comprendre la politique, l'économie et la culture. Grâce aux capacités de l'IA, le phénomène des données massives a eu un impact énorme, et probablement durable, sur la façon dont les individus et les groupes font des projets, obtiennent des informations sur eux-mêmes et sur le monde, se divertissent et se socialisent.

Les ordinateurs d'aujourd'hui sont technologiquement très puissants. Leurs algorithmes sont extrêmement sophistiqués. Leurs réseaux neuronaux reproduisent le traitement intellectuel des êtres humains et leur permettent de faire des analyses complexes. En traitant les données massives, les entreprises peuvent anticiper les choix et les préférences de leurs clients à un stade si précoce qu'elles peuvent prévoir ce que les clients veulent avant même qu'ils ne le sachent eux-mêmes. Grâce aux données massives, les processus d'entreprise passent d'une approche « réactive » à une approche « proactive »⁹³.

Internet joue un rôle fondamental dans ce scénario. Alors que les individus utilisent internet pour partager, pratiquement en continu, des informations, parfois sur eux-mêmes et leur vie, le web rassemble les matières premières à partir desquelles l'IA tirera des conclusions, fera des suppositions et trouvera des réponses aux questions posées. Le philosophe d'Oxford Luciano Floridi a inventé le concept de « onlife » pour décrire la

⁹² Pan S. B., « Get to know me: Protecting privacy and autonomy under big data's penetrating gaze », *Harvard Journal of Law and Technology* 30, 2016, p. 239,

<https://jolt.law.harvard.edu/assets/articlePDFs/v30/30HarvJLTech239.pdf> ; Surden H., « Artificial intelligence and law: An overview », *Georgia State University Law Review* 35, 2019, pp. 1311 et 1315.

⁹³ Microsoft Dynamics 365, *Delivering personalized experiences in times of change*, 2007, p. 3, <https://www.hso.com/wp-content/uploads/2020/03/Digitally-transforming-customer-experiences-ebook.pdf>.

transition fréquente et inconsciente des êtres humains entre le monde réel et le monde en ligne⁹⁴.

Ce phénomène s'intensifie. En 2023, on estime qu'il y aura plus de cinq milliards d'utilisateurs d'internet et 3,6 appareils par habitant, et que 70 % de la population mondiale disposera d'une connectivité mobile⁹⁵. Plus le monde sera connecté, plus la quantité de données massives produites sera importante. Ce n'est pas un hasard si l'une des questions les plus débattues actuellement est l'introduction des réseaux 5G, car ces derniers peuvent fournir un avantage informationnel considérable à leurs propriétaires.

Le secteur et l'industrie des médias jouent un rôle important dans ce scénario. Leur travail a toujours consisté à collecter, traiter et diffuser l'information. Grâce aux données massives, ils peuvent désormais faire le profil de leur public et savoir ce qu'il attend, comment présenter les actualités ou raconter une histoire, ou quelle serait la fin idéale pour un film. Les données massives permettent de personnaliser l'offre en identifiant les potentiels lecteurs d'informations ou les cinéphiles, car « les ordinateurs sont plus précis que les hommes pour tirer des prédictions en se fondant sur des "empreintes digitales", telles que des traits de caractère [ou] des attitudes politiques »⁹⁶.

La nouveauté apportée par les données massives modifie également le paysage médiatique. « ... [L]a télévision/les films/la musique numériques et une myriade de modèles de distribution en ligne mettent au défi les distributeurs en place (CD, câble) depuis des années... Les éditeurs en ligne exploitent les signaux des consommateurs à partir de ce qu'ils lisent, de l'endroit où ils se trouvent, des signaux sociaux qu'ils envoient – par exemple les articles qu'ils partagent, les sujets à la mode sur Facebook et Twitter – pour proposer un contenu personnalisé et pertinent tout en n'étant pas trop répétitif et prévisible, automatisant et surpassant ainsi ce que les éditeurs humains peuvent faire »⁹⁷. Lorsqu'il s'agit de produire des informations, les médias traditionnels sont désormais en concurrence avec des fournisseurs d'informations non professionnels qui fouillent le web à la recherche d'informations ou avec des blogueurs qui partagent leurs points de vue sur des plateformes de médias sociaux au sein desquelles la distribution et la consommation de contenu sont pratiquement indissociables⁹⁸.

Ce chapitre traite des répercussions juridiques les plus pertinentes qu'a un tel changement global sur le monde des médias. Il s'intéresse à la question cruciale de la

⁹⁴ Floridi L., « Soft ethics and the governance of the digital », *Philosophy & Technology* 31, 1, 2018, p. 1.

⁹⁵ Cisco, *Cisco Annual International Report (2018-2023) White Paper*, 9 mars 2020, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html?fbclid=IwAR31-e732ws1p1cIW5PYHQjVOJkPSzV0dGt3sq_qkX_P8wb9Q4Yn0Ez0a0Y.

⁹⁶ Contrôleur européen de la protection des données, Avis 7/2015 Relever les défis des données massives, 19 novembre 2015, p. 16, https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_fr.pdf.

⁹⁷ Byers A., « Big data, big economic impact », 10, 2015, https://kb.osu.edu/bitstream/handle/1811/75420/ISJLP_V10N3_757.pdf?sequence=1&isAllowed=y. Voir également Bruckner M. A., « The promise and perils of algorithmic lenders' use of big Data », *Chicago-Kent Law Review* 93, 2018, p. 8, <https://scholarship.kentlaw.iit.edu/cklawreview/vol93/iss1/1/> ou Ambrose M. L., « Lessons from the Avalanche of Numbers: Big Data in Historical Perspective », *ISJLP*, 11, 2015, p. 213, (« Netflix predicts our movies »).

⁹⁸ Perritt H. H. Jr., « Technologies of storytelling: New models for movies », *Virginia Sports & Entertainment Law Journal*, 10, 2010, p. 153, http://blogs.kentlaw.iit.edu/perrittseminar/files/2016/07/perritt-technologies-of-storytelling-Westlaw_Document_05_56_44.pdf.

protection de la vie privée. Il traite ensuite des possibles discriminations et biais auxquels une stratégie dirigée par les données massives peut se heurter et examine les risques de désinformation, de polarisation de la vie politique et de transformation du secteur des médias en un système de surveillance de masse. Plus loin, le chapitre présente une vue d'ensemble de l'évolution des marchés et des stratégies des médias à la lumière de la dynamique des données massives. Enfin, il réfléchit brièvement à une approche réglementaire qui serait appropriée en matière de données massives.

Au fil des ans, la nécessité de réglementer l'IA s'est majoritairement imposée. Bien que les technologies soient mondiales et ne connaissent pas de frontières, l'objectif, l'approche et le schéma réglementaires des grands acteurs juridiques dans ce scénario – les États-Unis, l'Union européenne et la Chine – divergent profondément. L'approche américaine vise à garantir que les marchés sur lesquels l'IA est massivement déployée restent ouverts et efficaces. La préoccupation première de l'UE semble consister à garantir le respect de la dignité de l'individu. Quant à la Chine, elle est surtout préoccupée par la paix sociale, la stabilité et le développement ordonné de son économie. Chacune de ces approches accorde aux données massives un traitement juridique spécifique.

2.2. Le respect de la vie privée comme gardien des données massives

De plus en plus de personnes craignent que les outils dirigés par les données massives ne s'intègrent dans un système omniprésent de surveillance et de manipulation de masse. L'une des principales garanties contre cette menace est la protection de la vie privée. De nombreux pays et systèmes juridiques supranationaux ont mis en place des réglementations qui limitent et contrôlent la nature et le mode de collecte et de traitement des informations, dans le but également de limiter les analyses des données massives et de prévenir les perturbations sociales. À cet égard, les lois sur la protection de la vie privée servent de bouclier contre les excès des données massives.

2.2.1. Les États-Unis d'Amérique

Le monde occidental est divisé quant à sa compréhension et sa protection de la vie privée. Les approches des États-Unis et de l'Union européenne sont loin d'être alignées. Malgré leur sensibilité historique au respect de la vie privée, les États-Unis ne disposent pas d'une réglementation complète sur la collecte et le rassemblement d'informations sur le web. Plusieurs régimes juridiques coexistent, chacun réglementant un secteur spécifique, sans qu'il y ait de réglementation globale à l'échelle nationale⁹⁹. L'approche américaine, cependant, considère généralement l'information comme un nouveau marché énorme,

⁹⁹ Houser K. A. & Voss W. G., « The end of Google and Facebook or a new paradigm in data privacy », *Richmond Journal of Law and Technology*, 25, 2018, p. 18, https://jolt.richmond.edu/files/2018/11/Houser_Voss-FE.pdf.

ayant des répercussions positives sur l'économie nationale. Si certains États, la Californie en tête, ont commencé à adopter des textes législatifs qui protègent et réglementent la vie privée, la collecte et le rassemblement de données à caractère personnel sont largement autorisés, voire encouragés. Une base juridique assez générale est que les personnes qui communiquent leurs données doivent simplement *savoir* que leurs informations seront traitées de diverses manières, y compris à des fins de profilage et d'exploitation commerciale de leurs goûts. Comme la plupart des acteurs de l'industrie mondiale de l'IA sont basés aux États-Unis, un schéma réglementaire aussi favorable leur permet d'exploiter pleinement les avantages du nouveau pétrole des données.

2.2.2. L'Union européenne

La protection de la vie privée au sein de l'Union européenne est basée sur le Règlement général sur la protection des données (RGPD)¹⁰⁰, adopté le 27 avril 2016 et entré en vigueur le 25 mai 2018. Le RGPD lui-même est le point culminant d'un long processus qui a renforcé la protection des données à caractère personnel au cours de plusieurs décennies, et représente un parcours très différent de celui des États-Unis. Bien que l'Union européenne se soit engagée à « faciliter l'accès des entreprises et des pouvoirs publics à des données de haute qualité pour stimuler la croissance et créer de la valeur »¹⁰¹, l'attitude générale de l'Union européenne repose sur le rejet de la marchandisation des données à caractère personnel¹⁰². La base juridique du RGPD est qu'un sujet doit donner son *consentement* au traitement de ses données¹⁰³. Le consentement lui-même doit être sans ambiguïté, donné librement et en toute connaissance de cause :¹⁰⁴ le sujet doit être informé de la portée et de la finalité du traitement¹⁰⁵. La protection du RGPD couvre les citoyens de l'UE ainsi que

¹⁰⁰ Texte consolidé : Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la Directive 95/46/CE (Règlement général sur la protection des données) (Texte présentant de l'intérêt pour l'EEE), <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:02016R0679-20160504>.

¹⁰¹ Contrôleur européen de la protection des données, Opinion 3/2020 on the European strategy for data, 16 juin 2020, p. 4, https://edps.europa.eu/sites/edp/files/publication/20-06-16_opinion_data_strategy_en.pdf. Voir également Conseil de l'Union européenne, Façonner l'avenir numérique de l'Europe - Conclusions du Conseil, 9 juin 2020, <https://data.consilium.europa.eu/doc/document/ST-8711-2020-INIT/fr/pdf>.

¹⁰² Comité européen de la protection des données, Lignes directrices 2/2019 sur le traitement des données à caractère personnel au titre de l'article 6, paragraphe 1, point b), du RGPD dans le cadre de la fourniture de services en ligne aux personnes concernées, Version 2.0, 8 octobre 2019, n° 54, https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_fr.pdf.

¹⁰³ Article 6 RGPD.

¹⁰⁴ Manheim K. & Kaplan L., « Artificial intelligence: Risks to privacy and democracy », *Yale Journal of Law & Technology*, 106, 2019, p. 1069, https://yolt.org/sites/default/files/21_yale_j.l._tech._106_0.pdf.

¹⁰⁵ Article 6, paragraphe 4, et 7, RGPD.

les données de toute autre personne physique, pour autant que le traitement ait lieu au sein de l'UE. En d'autres termes, elle protège toute personne se trouvant sur son territoire¹⁰⁶.

La différence entre les approches américaine et européenne a créé une faille dans l'échange de données de part et d'autre de l'Atlantique. Le RGPD est très prudent quant au partage des informations recueillies au sein de l'Union européenne, et exige que tout transfert de données en dehors des frontières de l'UE soit conforme aux normes de l'UE¹⁰⁷. La philosophie réglementaire de l'UE a été perçue comme étant si protectrice de la vie privée que de nombreux citoyens non européens ont tendance à préférer les entreprises basées dans l'UE aux entités qui ne sont pas soumises à la juridiction de l'Union européenne. La conformité au RGPD est donc devenue un atout de réputation pour les entreprises travaillant dans le domaine de l'IA même en dehors de l'Union européenne, les poussant à mettre en œuvre spontanément des règles de protection de la vie privée¹⁰⁸.

Un tel niveau de protection de la vie privée par le RGPD a cependant un coût. La nécessité pour les entreprises d'obtenir le consentement des internautes qui visitent leurs sites web se traduit par une pléthore de demandes de consentement répétitives, et parfois obscures, qui apparaissent traditionnellement dès qu'une page web s'affiche¹⁰⁹. Ce phénomène a inondé internet au point que la plupart des utilisateurs cliquent simplement sur « oui » et continuent de naviguer sur le site sans prêter attention à la manière dont leurs informations sont collectées, traitées et diffusées¹¹⁰. Cette ligne de conduite est certes risquée mais compréhensible. Il a été estimé qu'une personne normale – pas un avocat compétent ou un internaute excessivement méticuleux – perdrait 76 jours par an à lire tous les avertissements relatifs à la vie privée qui apparaissent lorsqu'il est en ligne¹¹¹. Une protection excessive de la vie privée peut s'avérer contre-productive : les personnes peuvent renoncer à toute la protection en donnant leur consentement de manière trop superficielle, permettant ainsi une collecte massive de leurs informations.

De plus, les possibilités d'analyse de données massives peuvent affaiblir la protection de la vie privée accordée par le RGPD sur de nombreux fronts. Premièrement, le RGPD impose moins de restrictions aux données anonymes, car l'anonymat est censé protéger la vie privée. Toutefois, grâce aux capacités croissantes de l'IA, « il est de plus en plus facile de déduire l'identité d'une personne de la combinaison de données prétendument "anonymes" et d'autres ensembles de données, y compris d'informations à la

¹⁰⁶ Contrôleur européen de la protection des données, Avis 3/2018 Avis du CEPD sur la manipulation en ligne et les données à caractère personnel, 19 mars 2018, p. 14, https://edps.europa.eu/sites/edp/files/publication/18-03-19_opinion_online_manipulation_fr.pdf.

¹⁰⁷ Article 45 RGPD.

¹⁰⁸ Moerel L. & Lyon C., « Commoditization of data is the problem, not the solution – Why placing a price tag on personal information may harm rather than protect consumer privacy », *Future of Privacy Forum*, 24 juin 2020, <https://fpf.org/2020/06/24/commoditization-of-data-is-the-problem-not-the-solution-why-placing-a-price-tag-on-personal-information-may-harm-rather-than-protect-consumer-privacy>.

¹⁰⁹ Contrôleur européen de la protection des données, Avis 7/2015 Relever les défis des données massives, *op. cit.*, p. 11.

¹¹⁰ Tsesis A., « Marketplace of ideas, privacy, and the digital audience », *Notre Dame Law Review*, 94, 2019, p. 1590, <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4845&context=ndlr>.

¹¹¹ Hartzog W., *Privacy's blueprint*, Harvard University Press, 2018.

disposition du public, par exemple sur les médias sociaux »¹¹²... « Plus une collecte de données est importante et exhaustive », plus il existe de possibilités qu'une personne dont les données ont été anonymisées soit réidentifiée¹¹³.

En outre, les règles de l'UE en matière de respect de la vie privée exigent que les personnes reçoivent des informations détaillées sur la finalité et l'étendue du traitement des données qu'elles confèrent. Grâce aux réseaux neuronaux et à l'apprentissage profond, les systèmes basés sur l'IA tirent des conclusions que même les développeurs de logiciel ne peuvent pas totalement anticiper. Cette capacité même des données massives compromet l'interprétation de la réglementation de l'UE en matière de protection de la vie privée. Comme le traitement des données massives donne des résultats qui ne peuvent être entièrement prévus, il est extrêmement difficile de donner aux personnes une idée détaillée de l'usage qui sera fait de leurs informations¹¹⁴.

2.2.3. La Chine

Les institutions publiques et privées chinoises tirent des quantités massives de données d'une multitude de sources pour établir le profil des individus avec le plus haut degré de précision. La collecte et le traitement de données à caractère personnel sur la population chinoise sont essentiels au grand plan civique chinois, qui prévoit la mise en œuvre d'un vaste programme de surveillance et de suivi qui exploite l'IA pour établir le profil et prévoir les comportements des individus et des groupes¹¹⁵. L'objectif global de ce plan consiste à construire un système de crédit social omniprésent – un mécanisme basé sur l'IA qui recueille des informations à partir des dossiers personnels, des smartphones et des systèmes de surveillance de masse, puis classe les individus et leur accorde des privilèges et des droits en fonction de leur conduite antérieure¹¹⁶.

En Chine, les institutions publiques s'efforcent de rendre la vie de chacun transparente, et non pas privée, et s'associent à des entreprises privées chinoises à cette

¹¹² Contrôleur européen de la protection des données, Avis 4/2015. Vers une nouvelle éthique numérique, 11 septembre 2015, p. 6, https://edps.europa.eu/sites/edp/files/publication/15-09-11_data_ethics_fr.pdf.

¹¹³ Contrôleur européen de la protection des données, Avis 7/2015, « Relever les défis des données massives », *op. cit.*, p. 15.

¹¹⁴ AGCM, AGCOM, et Garante per la protezione dei dati personali, Indagine conoscitiva sui *Big Data*, pp. 25-26, <https://www.agcom.it/documents/10179/17633816/Documento+generico+10-02-2020+1581346981452/39c08bbe-1c02-43dc-bb8e-6d1cc9ec0fcf?version=1.0>. Le document explique comment le « consentement dynamique » est en train de devenir une option viable dans le cadre du système réglementaire de l'UE en matière de protection de la vie privée. Ce concept comprend le consentement comme un processus progressif, au cours duquel le sujet peut être contacté plusieurs fois pour lui demander s'il consent à une utilisation donnée de ses informations.

¹¹⁵ Conseil d'État, Avis du Conseil d'État publiant le plan de développement de la nouvelle génération de l'intelligence artificielle, n° 358 juillet 2017, pp. 2-5 et 18-21, <https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence-development-plan>.

¹¹⁶ Conseil d'État, Avis concernant la publication du plan de planification pour la construction d'un système de crédit social (2014-2020), n° 21, 14 juin 2014, <https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020>.

fin. Une poignée de grandes entreprises technologiques telles que WeChat et Alibaba fonctionnent ainsi comme des plaques tournantes numériques dans la vie des citoyens chinois¹¹⁷. Les Chinois sont encouragés à utiliser la même application mobile pour un large éventail d'activités, de la réservation d'un taxi au paiement d'un restaurant, en passant par les relations sociales ou les interactions avec l'administration publique. Une énorme quantité d'informations sur toute personne est ainsi recueillie et transmise aux institutions publiques à des fins de profilage¹¹⁸.

2.2.4. Trois approches différentes ?

Aussi étrange que cela puisse paraître, certains soutiennent qu'un système de crédit social similaire est déjà en place dans le secteur privé des États-Unis¹¹⁹. Les entreprises privées ne se contentent pas de profiler leurs clients pour les fidéliser. Elles vendent également les informations les concernant à d'autres entreprises. Les préférences personnelles et les habitudes d'achat sont ainsi mises en adéquation pour mieux profiler les utilisateurs, anticiper leurs décisions et orienter leurs comportements¹²⁰. Une banque ou une compagnie d'assurances peut évaluer avec précision le risque financier d'un individu sur la base de diverses informations, allant de son éducation, de son mode de vie ou des lieux qu'il fréquente et des personnes auxquelles il rend visite. Un parti politique peut évaluer la sensibilité politique d'un individu en fonction des films qu'il regarde, des chaînes de médias qu'il préfère ou de ses antécédents familiaux.

Il est peu surprenant, voire pas du tout, que l'approche globale des États-Unis en matière de protection des données néglige le potentiel négatif d'une telle accumulation privée de données à caractère personnel. La culture américaine des droits a toujours cherché à conserver le contrôle sur les pouvoirs publics. Cette approche reste d'actualité et maintient l'attention des États-Unis sur les menaces que constituent les pouvoirs publics, alors que l'Europe a toujours été plus attentive à la capacité des entreprises privées à enfreindre les droits fondamentaux¹²¹. Le résultat paradoxal est que les États-Unis sont la

¹¹⁷ Pieranni S., *Red Mirror*, Laterza, 2020, pp. 22-23.

¹¹⁸ *Ibid*, pp. 40 et 115.

¹¹⁹ Baker L. C., « Next generation law: Data-driven governance and accountability-based regulatory systems in the West, and social credit regimes in China », *Southern California Interdisciplinary Law Journal*, 28, 2018, pp. 170-171, <https://lbackerblog.blogspot.com/2019/05/just-published-next-generation-law-data.html>.

¹²⁰ Le Parlement européen a récemment invité la Commission européenne à « interdire aux plateformes d'afficher des publicités microciblées » : Parlement européen, Résolution du 18 juin 2020 sur la politique de concurrence – rapport annuel 2019, https://www.europarl.europa.eu/doceo/document/TA-9-2020-0158_FR.html. Selon Morozov E., « Digital socialism? », *New Left Review*, 116/117, mars-juin 2019, p. 62, <https://newleftreview.org/issues/II116/articles/evgeny-morozov-digital-socialism>, « Amazon a obtenu un brevet sur l'«expédition prédictive», ce qui lui permet de nous envoyer des produits avant même que nous sachions que nous les voulons ».

¹²¹ En ce qui concerne la vigilance européenne face au potentiel nuisible des entreprises privées, voir le contrôleur européen de la protection des données, Avis 8/2016 Avis du CEPD sur une application cohérente des droits fondamentaux à l'ère des données massives (*Big Data*), 23 septembre 2016, p. 5, https://edps.europa.eu/sites/edp/files/publication/16-09-23_bigdata_opinion_fr.pdf. Voir également Pollicino

plaque tournante mondiale de l'innovation en matière de données massives, mais ne voient pas la menace que constituent ces données pour les droits fondamentaux comme l'Europe semble le faire.

Ces approches différentes de la vie privée ont de puissantes conséquences sur la vie ordinaire des citoyens et des entreprises de médias. Comme nous le verrons plus loin, l'exploitation des technologies basées sur l'IA transforme les groupes de médias en bien plus que des fournisseurs d'informations ; ils peuvent devenir des collecteurs d'informations et participer au profilage des individus.

2.3. Biais et discrimination des données massives

Bien qu'on ne s'attende pas à ce qu'un logiciel soit biaisé, l'un des plus grands défis pour les technologies dirigées par les données est leur potentiel discriminatoire. La collecte d'informations, le traitement et la diffusion de l'information peuvent incorporer, intégrer et amplifier les préjugés. L'exemple le plus célèbre est probablement celui du *chatbot* (robot conversationnel) Tay de Microsoft. En 2016, Microsoft a créé un profil Facebook pour un logiciel innovant capable d'interagir sur la plateforme médiatique avec d'autres utilisateurs de Facebook en recueillant des informations sur le web, en identifiant les tendances et en échangeant des opinions en conséquence.

En l'espace de 16 heures, le compte Facebook a été ouvert puis fermé, après que ses créateurs se sont rendu compte que Tay tenait des propos sexistes et racistes¹²². Les développeurs du logiciel n'ont certainement pas fourni à leur robot tous les préjugés qu'il a ensuite affichés sur le web. Ses créateurs ont simplement utilisé le web comme base d'apprentissage pour le robot, lequel a manifestement trouvé que le racisme et le sexisme étaient des comportements répandus qui attireraient l'attention. Tay a façonné son langage et ses thèmes en fonction de la formation qu'il a suivie. Il a appris et adopté des préjugés par lui-même.

La vie éphémère de Tay explique l'importance de la formation pour l'IA. Les systèmes basés sur l'IA ont besoin de beaucoup de données pour apprendre. Plus ils recueillent d'informations, plus ils sont capables de faire des déductions et des choix. Malheureusement, les grands ensembles de données pour former les algorithmes ne sont pas souvent disponibles, de sorte que les programmeurs de logiciel exploitent souvent ce qui est déjà disponible sur le web. Ce choix est extrêmement problématique car les êtres humains ne peuvent pas superviser entièrement le processus d'apprentissage, et l'IA peut

O., « L'«autunno caldo» della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale », *Federalismi*, 15 octobre 2019, <https://www.federalismi.it/nv14/editoriale.cfm?eid=533>.

¹²² « Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot », *The Guardian*, 26 mars 2016, <https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot>.

prendre des directions imprévues, voire inappropriées. Elle peut identifier et intégrer les biais de la société, et les stimuler par son activité¹²³.

Des ensembles de données déséquilibrés peuvent involontairement créer des biais, comme le montre le cas de la reconnaissance faciale. Les systèmes d'IA occidentaux de reconnaissance faciale ne parviennent souvent pas à identifier correctement les individus qui ne sont pas de type caucasien parce que les autres groupes ethniques apparaissent moins souvent sur le web que ce type, tandis que les logiciels d'IA développés en Chine souffrent du problème inverse¹²⁴. Par conséquent, il est plus probable, que dans les pays occidentaux par exemple, une personne noire soit prise pour quelqu'un d'autre, qu'un Caucasien ne le soit. Les systèmes médiatiques qui intègrent des processus basés sur les données massives sont donc confrontés à un véritable défi, car en exploitant l'IA, ils peuvent intégrer des préjugés et des déséquilibres sociaux.

Il est très difficile de lutter contre la discrimination dans le domaine des données massives et des réseaux neuronaux en raison des dangers de la « discrimination par procuration »¹²⁵. La discrimination par procuration est une politique privée ou publique qui comprend une condition ou un facteur apparemment neutre mais intégrant en réalité une tradition, une pratique ou une croyance discriminatoire. Par exemple, dans les sociétés socialement ou territorialement divisées, le code postal ou le prix du logement peut servir de discrimination par procuration pour les polices d'assurance ou le zonage, car il peut défavoriser certains groupes ethniques et en favoriser d'autres. Même si les développeurs de logiciel interdisent expressément à l'IA de prendre en compte l'appartenance ethnique à des fins de déductions, d'autres facteurs peuvent servir de relais à la discrimination¹²⁶. Dans une société donnée, des stratégies de marché, des campagnes politiques ou des fournisseurs de services sociaux dirigés par les données massives peuvent – même involontairement – isoler et discriminer systématiquement les groupes les plus défavorisés par procuration.

2.4. Informer le public : médias, désinformation et contenu illégal

L'IA est un outil médiatique puissant. Elle peut découvrir des faits, détecter les goûts, établir le profil des utilisateurs et anticiper les tendances sociales. En quelques mots, elle peut fournir aux gens davantage que ce qu'ils veulent recevoir. Mais la personnalisation des offres médiatiques par le biais de données massives a un prix.

¹²³ Stevenson M. T. & Doleac J. L., *Algorithmic Risk Assessment in the Hands of Humans*, Institute of Labor Economics, 1^{er} décembre 2019, p. 1, <http://ftp.iza.org/dp12853.pdf> ; Bruckner M. A., *op. cit.*, p. 25.

¹²⁴ Grother P., Ngan M., Hanaoka K., « Face recognition vendor test (FRVT) Part III. Demographic effects », National Institute of Standards and Technology Interagency 8280, décembre 2019, <https://doi.org/10.6028/NIST.IR.8280>.

¹²⁵ Prince A. E. R. & Schwarcz D., « Proxy discrimination in the age of artificial intelligence and big data » *Iowa Law Review* 105, 2020, p. 1260, <https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data>.

¹²⁶ Idem.

L'IA est un excellent outil pour présélectionner des contenus susceptibles d'intéresser les utilisateurs de médias. Compte tenu du trop-plein d'informations, la capacité de l'IA à établir le profil d'un utilisateur permet de prévoir ses intérêts pour une information, ce qui rend le travail des médias plus efficace et l'expérience de l'utilisateur plus agréable. Cependant, les utilisateurs des médias peuvent ne pas se rendre compte que leur horizon se rétrécit – que le type d'informations qu'ils reçoivent peut ne pas représenter la réalité avec exactitude, mais seulement la « réalité » de ce que l'IA comprend comme reflétant leurs intérêts.

Nourrir les utilisateurs en leur donnant encore plus de ce qu'ils préfèrent, connaissent ou de ce à quoi ils s'intéressent déjà, tend à créer des bulles sociales. Les technologies des données massives peuvent filtrer l'information en fonction de ce qu'un utilisateur de médias est censé aimer ou croire. Au lieu d'élargir l'horizon des utilisateurs, l'IA est ainsi en mesure de stimuler la sélectivité intellectuelle des individus. Une industrie de l'information conviviale peut perdre de vue son objectif, à savoir fournir à la société de larges perspectives, des informations pleinement avisées et des points de vue stimulants.

Les stratégies médiatiques reposant sur les données massives peuvent ainsi déclencher involontairement la création de bulles informationnelles. Il existe toutefois le risque supplémentaire qu'une bulle soit générée intentionnellement. Les grandes entreprises technologiques peuvent profiler des utilisateurs et des informations afin de stimuler ou d'entraver la diffusion de certaines informations en fonction de leurs stratégies de marché ou de leurs priorités¹²⁷.

Les données massives opposent également les médias traditionnels aux médias sociaux. Les médias sociaux exploitent la forte protection normalement accordée à la liberté d'expression, et vivent de leur présence continue sur le web et de leur capacité à alimenter le public en informations¹²⁸. Bon marché et facilement accessibles, ils peuvent ainsi remplacer les opérateurs et points de vente professionnels de médias. Cette concurrence asymétrique a déclenché un dangereux « nivellement par le bas » dans le domaine des fournisseurs d'informations¹²⁹. Pour éviter de perdre leur audience, les médias traditionnels essaient de suivre la cadence des services non professionnels tels que les blogs, souvent au détriment de l'exactitude¹³⁰.

Les bulles des plateformes de médias basées sur l'IA participent souvent à la diffusion de fausses nouvelles. Fléau dans l'industrie de l'information d'aujourd'hui, selon certaines statistiques, les fausses nouvelles sont capables d'atteindre plus de gens et plus rapidement que les informations vérifiées et contrôlées¹³¹, donnant vie à ce que Cass

¹²⁷ Singer H., « How Washington should regulate Facebook », *Forbes*, 18 octobre 2017, <https://www.forbes.com/sites/washingtonbytes/2017/10/18/what-to-do-about-facebook>.

¹²⁸ Shefa M. C., « First Amendment 2.0: Revisiting Marsh and the quasi-public forum in the age of social media », *University of Hawaii Law Review*, 41, 2018, p. 160.

¹²⁹ AGCM, AGCOM, et Garante per la protezione dei dati personali, Indagine conoscitiva sui Big Data, *op. cit.*, p. 30.

¹³⁰ Contrôleur européen de la protection des données, Avis 3/2018 Avis du CEPD sur la manipulation en ligne et les données à caractère personnel, *op. cit.* p. 13 (« Certaines preuves montrent que... concentration et élimination du journalisme local facilite la propagation de la désinformation »).

¹³¹ Idem.

Sunstein a appelé des « cybercascades »¹³². Le système des bulles aggrave le processus, car il filtre les faits et les différents points de vue, renforçant ainsi les points de vue profondément ancrés, voire les préjugés.

Les stratégies dirigées par les données massives remettent en question le rôle historique joué par le système des médias et la liberté d'expression dans les régimes démocratiques. Au lieu d'élargir les horizons, de remettre en question les points de vue, de mettre à nu les biais et de faire progresser la société, les plateformes médiatiques contemporaines risquent d'isoler mutuellement les groupes sociaux et de renforcer les opinions profondément ancrées. Traditionnellement, le constitutionnalisme libéral valorise et protège grandement la liberté d'expression car les différents points de vue font progresser les sociétés grâce au libre échange d'opinions. Au contraire, les technologies des données massives sont capables de créer des « chambres d'écho »¹³³, qui expulsent la dissidence et gravitent autour de croyances non contestées. Les opinions qui remettent en question des visions du monde profondément ancrées sont éjectées d'une bulle et trouveront probablement leur place dans une autre bulle qui n'offre pratiquement aucun échange en dehors d'elle-même.¹³⁴ Les données massives peuvent donc restreindre les perspectives et immuniser les préjugés contre les avantages de la liberté d'expression.

Les institutions privées et publiques ont pris conscience des distorsions que les données massives peuvent causer aux médias et à la société en général. Par exemple, Twitter a récemment créé un outil de vérification des faits litigieux dans le but de détecter les fausses nouvelles ou les tweets qui nuisent à des groupes identifiables¹³⁵. Le *Code européen de bonnes pratiques contre la désinformation*¹³⁶ préconise un examen approfondi du phénomène, en soulignant que « toutes les parties prenantes ont un rôle à jouer dans la lutte contre la propagation de la désinformation ». Les signataires du code, parmi lesquels figurent Facebook, Google, Mozilla, TikTok et Twitter, ont ainsi promis de « réduire la visibilité de la désinformation en améliorant la possibilité de trouver des contenus fiables », et de « faciliter la découverte de contenus et l'accès à différentes sources d'information représentant des points de vue alternatifs ». Nombreux sont ceux qui demandent la réglementation du déploiement de l'IA de sorte à soumettre les fournisseurs d'accès internet à des règles similaires aux « normes traditionnelles relatives à la responsabilité des médias »¹³⁷.

¹³² Sunstein C. R., « *#republic: Divided democracy in the age of Social Media* », Princeton University Press, 2017, p. 57.

¹³³ Sasahara K. *et al.*, « On the inevitability of online echo chambers », <https://arxiv.org/abs/1905.03919>.

¹³⁴ Jones R. L., « Can you have too much of a good thing: The modern marketplace of ideas », *Missouri Law Review*, 83, 2018, p. 987, <https://scholarship.law.missouri.edu/mlr/vol83/iss4/8/>.

¹³⁵ Pham S., « Twitter says it labels tweets to provide 'context, not fact-checking' », *CNN Business*, <https://edition.cnn.com/2020/06/03/tech/twitter-enforcement-policy/index.html>.

¹³⁶ Code européen de bonnes pratiques contre la désinformation, <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>.

¹³⁷ Contrôleur européen de la protection des données, Avis 3/2018 Avis du CEPD sur la manipulation en ligne et les données à caractère personnel, *op. cit.*, p. 16.

Les politiques de l'UE visent en particulier les contenus terroristes, les contenus pédopornographiques, le racisme, les discours xénophobes et haineux¹³⁸, qui sont généralement des sujets de préoccupation majeurs pour les médias sociaux d'aujourd'hui. En fait, étant donné l'afflux massif de données, il est techniquement irréaliste de filtrer les informations avant de décider de les héberger. Les plateformes en ligne combinent donc normalement deux systèmes distincts : d'une part, elles adoptent un système de « notification et retrait » – toute personne peut se plaindre qu'un affichage spécifique de contenu est en infraction avec la loi et demander à la plateforme médiatique de procéder à une évaluation ; d'autre part, la plupart des plateformes adoptent des systèmes de filtrage basés sur les données massives qui passent au crible les contenus de façon automatique et continue, en prenant des décisions sur ce qui doit être caché au public¹³⁹. La plupart des plateformes disposent d'une protection supplémentaire contre ces décisions automatisées, qui permet aux individus de contester la décision prise par un logiciel de supprimer certains matériels¹⁴⁰.

Aux États-Unis et dans l'Union européenne qui dispose de « l'un des cadres réglementaires les plus complets pour le suivi des contenus illégaux en ligne »¹⁴¹, les fournisseurs de services bénéficient de larges exemptions de responsabilité. Ces exemptions visent à préserver leur rôle positif dans l'établissement de liens entre les personnes et la diffusion de l'information¹⁴². Le droit de l'UE a renforcé cette règle en interdisant à ses États membres d'imposer aux plateformes d'hébergement des obligations générales de surveillance du matériel qu'elles hébergent¹⁴³. Le scénario est cependant en

¹³⁸ Département politique des politiques économiques, scientifiques et de la qualité de la vie, « Online platforms' moderation of illegal content online », juin 2020, p. 9,

[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU\(2020\)652718_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652718/IPOL_STU(2020)652718_EN.pdf).

¹³⁹ *Ibid*, p. 45.

¹⁴⁰ *Ibid*, p. 10.

¹⁴¹ *Ibid*, p. 66.

¹⁴² Pour les États-Unis, voir le titre 47, section 230 de la Communication Decency Act, <https://www.fcc.gov/general/telecommunications-act-1996> ; pour l'UE, voir la 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 »),

<https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32000L0031&from=FR>, art. 14 : « 1. Les États membres veillent à ce que, en cas de fourniture d'un service de la société de l'information consistant à stocker des informations fournies par un destinataire du service, le prestataire ne soit pas responsable des informations stockées à la demande d'un destinataire du service à condition que: a) le prestataire n'ait pas effectivement connaissance de l'activité ou de l'information illicites et, en ce qui concerne une demande en dommages et intérêts, n'ait pas connaissance de faits ou de circonstances selon lesquels l'activité ou l'information illicite est apparente ou b) le prestataire, dès le moment où il a de telles connaissances, agisse promptement pour retirer les informations ou rendre l'accès à celles-ci impossible ». Quant à la protection des mineurs, voir Directive 2010/13/UE du Parlement européen et du Conseil du 10 mars 2010 visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à la fourniture de services de médias audiovisuels (Directive Services de médias audiovisuels) (version codifiée) (texte présentant de l'intérêt pour l'EEE). Une version consolidée incluant les modifications introduites en 2018 est disponible sur :

<https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:02010L0013-20181218>.

¹⁴³ Département politique des politiques économiques, scientifiques et de la qualité de la vie, *op. cit.*, p. 21.

évolution¹⁴⁴. Dans son interprétation de la directive sur le commerce électronique, la Cour de justice de l'Union européenne a estimé que les prestataires de services qui ne se contentent pas d'afficher passivement des matériels sont censés faire plus que simplement examiner et supprimer ces matériels lorsque cela est nécessaire, après qu'on leur a demandé de le faire¹⁴⁵. En fait, la Cour a estimé qu'une décision judiciaire de suppression s'étend « aux informations dont le contenu, tout en véhiculant en substance le même message [auquel la décision judiciaire se réfère], est formulé de manière légèrement différente, en raison des mots employés ou de leur combinaison, par rapport à l'information dont le contenu a été déclaré illicite »¹⁴⁶. Certains ont critiqué ce principe sensé car il aboutirait au « paradoxe du bon samaritain » : plus une plateforme s'engage à surveiller les informations qu'elle publie, plus elle devient responsable. On craint qu'une telle approche judiciaire n'encourage les prestataires à rester passifs et à limiter leur activité de surveillance afin d'éviter les risques de responsabilité¹⁴⁷. Il s'agit maintenant de savoir si l'UE doit revoir sa politique et imiter l'approche américaine, qui a consisté à préserver l'exemption de responsabilité pour les plateformes, car cela les encouragerait à devenir plus proactives, ou si cela compromettrait la protection des individus et des groupes¹⁴⁸.

Dans le contexte des documents illégaux publiés sur les plateformes en ligne, l'IA peut certainement jouer un rôle important. Compte tenu de l'énorme quantité de données échangées et de la tendance à créer des bulles dans lesquelles les utilisateurs des médias ne trouvent guère d'informations qu'ils n'aiment pas ou de points de vue avec lesquels ils sont en désaccord, les matériels illégaux peuvent ne pas être détectés par les êtres humains pendant longtemps. Le développement de systèmes basés sur l'IA qui filtrent le contenu peut donc devenir souhaitable, voire nécessaire. L'IA et les données massives ne sont pas seulement une partie du problème – elles peuvent faire partie de la solution. Il est évident que le contrôle basé sur l'IA ne doit pas devenir une forme de censure automatisée. Les fournisseurs pourraient exploiter les systèmes d'IA pour filtrer les matériels qui sont simplement controversés, isolant ainsi la sphère publique des opinions ou des informations minoritaires au sujet desquelles beaucoup auraient du mal à nouer le dialogue. Ce risque doit être maîtrisé.

¹⁴⁴ Nunziato D. C., « The marketplace of ideas online », *Notre Dame Law Review*, 94, 2019, p. 1521, <https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=4844&context=ndlr>.

¹⁴⁵ C-324/09, *L'Oréal et al. contre eBay International AG*, paragraphes 113-115, <http://curia.europa.eu/juris/document/document.jsf?text=&docid=107261&pageIndex=0&doclang=fr&mode=lst&dir=&occ=first&part=1&cid=12642628>.

¹⁴⁶ C-18/18, *Eva Glawischnig-Piesczek contre Facebook Ireland Ltd.*, paragraphe 41, <http://curia.europa.eu/juris/document/document.jsf?text=&docid=218621&pageIndex=0&doclang=fr&mode=lst&dir=&occ=first&part=1&cid=12642666>.

¹⁴⁷ Département politique des politiques économiques, scientifiques et de la qualité de la vie, *op. cit.* p. 20 ; Policy Department Economic and Scientific Policy, « Liability of Online Service Providers for Copyrighted Content - Regulatory Action Needed », janvier 2018, p. 10, [https://www.europarl.europa.eu/RegData/etudes/IDAN/2017/614207/IPOL_IDA\(2017\)614207_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2017/614207/IPOL_IDA(2017)614207_EN.pdf).

¹⁴⁸ *Ibid*, p. 67.

2.5. La politique des données massives et la bulle politique¹⁴⁹

Les démocraties ont besoin d'une sphère publique solide pour survivre et s'épanouir¹⁵⁰. L'existence et l'échange de visions du monde et d'opinions politiques alternatives sont indispensables à leur survie. Plus généralement, au sein des démocraties, « les gens devraient être exposés à des matériels qu'ils n'auraient pas choisis à l'avance »¹⁵¹, car l'un des avantages historiquement associé aux démocraties est que « les biais sont filtrés dans la vaste république »¹⁵².

Les médias sociaux ont inondé la politique contemporaine. Les facultés de droit et les tribunaux ont réagi en développant lentement mais sûrement l'idée classique des forums publics pour y intégrer également les sites de médias sociaux privés¹⁵³. En raison de leur rôle social omniprésent et de leur importance cruciale pour fournir au public des flux d'informations et des opinions politiques, la Cour suprême des États-Unis a qualifié les sites de médias sociaux de « place publique moderne »¹⁵⁴. Ils sont tellement essentiels à la vie sociale et politique, a fait valoir la Cour, qu'ils doivent être accessibles au grand public¹⁵⁵. Depuis 2001, les tribunaux américains ont également « considéré les ordinateurs et l'accès à internet comme “pratiquement indispensables dans le monde moderne des communications et de la collecte d'informations” »¹⁵⁶.

Les médias sociaux ne sont cependant pas des lieux universellement accessibles dans lesquels tout le monde est accueilli et peut faire valoir ses arguments. L'analyse des données massives permet aux médias sociaux de segmenter la sphère publique en bulles autoréférentielles¹⁵⁷. Même les plateformes médiatiques qui ne filtrent pas intentionnellement l'information adaptent leurs flux d'informations aux besoins et aux choix de leurs utilisateurs, créant ainsi des bulles informationnelles. De telles bulles sont capables de diviser l'opinion publique en sphères d'influence homogènes et impénétrables¹⁵⁸.

La création de chambres d'écho homogènes, partisans et non conversationnelles ne remplace pas le pluralisme¹⁵⁹ démocratique et peut même le menacer¹⁶⁰. Le scandale de Cambridge Analytica, qui aurait récolté les données des utilisateurs de Facebook sans leur consentement afin de développer des « profils psychographiques » avant de cibler des

¹⁴⁹ Pour un point de vue différent sur la question des bulles de filtres/chambres d'écho, voir le chapitre 5 de cette publication.

¹⁵⁰ Wischmeyer T., « Making social media an instrument of democracy », *European Law Journal*, 25, 2019, p. 172, <https://onlinelibrary.wiley.com/doi/abs/10.1111/eulj.12312>.

¹⁵¹ Sunstein C. R., *op. cit.* p. 6.

¹⁵² McGinnis J. O., *Accelerating Democracy*, Princeton University Press, 2013, p. 127.

¹⁵³ Nunziato D. C., *op. cit.*, p. 3.

¹⁵⁴ *Packingham v. North Carolina* 582 U.S. ____ (2017), https://www.supremecourt.gov/opinions/16pdf/15-1194_0811.pdf.

¹⁵⁵ *Ibid.*

¹⁵⁶ Shefa M. C., *op. cit.*, p. 164.

¹⁵⁷ Sunstein C. R., *op. cit.*

¹⁵⁸ Sasahara K. *et al.*, *op. cit.*

¹⁵⁹ Wischmeyer T., *op. cit.*, pp. 173-174.

¹⁶⁰ Manheim K. & Kaplan L., *op. cit.* p. 109.

individus sélectionnés pour orienter leurs comportements de vote¹⁶¹, n'est qu'un exemple de la façon dont les données massives peuvent affecter la vie politique¹⁶². Et il est prouvé que des robots alimentés par les données massives sont déployés pour influencer les agendas politiques¹⁶³.

Cass Sunstein, professeur de droit à Harvard, a étudié l'impact des plateformes de médias sociaux basées sur l'IA dans la sphère politique dans son ouvrage célèbre, *#Republic*¹⁶⁴. Sunstein a démontré de manière convaincante la capacité de l'IA à générer des groupes d'information et à polariser la vie politique. Les campagnes politiques peuvent cibler des utilisateurs bien profilés, les exposant à certaines opinions ou certains faits tout en faisant taire ou en minimisant les déclarations des opposants politiques ou les faits qui remettraient en question leur propre programme et leur propre agenda¹⁶⁵. L'IA divise ainsi la sphère publique en environnements homogènes qui n'interagissent pratiquement pas entre eux. Les politiciens qui réussissent vont souvent à l'extrême pour galvaniser leurs partisans et renforcer le système des bulles.

La politique des données massives brouille souvent la ligne entre capacité personnelle et capacité institutionnelle. De nombreuses personnalités politiques préfèrent utiliser leur profil personnel dans les médias sociaux plutôt que leur profil institutionnel pour également communiquer avec le grand public sur des questions institutionnelles. En utilisant leurs profils personnels, ils forcent le public – qui suivrait normalement les pages et les profils des médias institutionnels – à entrer dans leur sphère de partisans.

Certains systèmes juridiques ont déployé des contre-mesures pour lutter contre cette privatisation de la sphère publique en chambres d'écho médiatiques séparées. L'expérience américaine est l'exemple le plus éloquent de cette évolution. De nombreuses personnalités publiques – dont le président Donald Trump – qui ont utilisé des sites web personnels à des fins institutionnelles ont bloqué les personnes qui faisaient des commentaires critiques sur leurs messages (ou « posts »), les empêchant ainsi de s'exprimer devant les citoyens¹⁶⁶. Certains citoyens ainsi expulsés du public ont poursuivi les politiciens concernés – et ont obtenu gain de cause devant les tribunaux. Les juges ont examiné la structure des plateformes médiatiques et la manière dont les politiciens les utilisent, et ont conclu que ces plateformes devaient être considérées comme des lieux publics qui devraient rester ouverts à tous. Les hommes politiques pouvaient toujours « mettre en sourdine » leurs *followers*, les empêchant ainsi d'engager une conversation dans le cadre de

¹⁶¹ *Ibid*, p. 139.

¹⁶² Pour d'autres exemples provenant de divers pays, voir Gurumurthy A. et Bharthur D., « Democracy and the algorithmic turn », *Sur International Journal of Human Rights*, 27, 2018, pp. 43-44, <https://sur.conectas.org/en/democracy-and-the-algorithmic-turn>, et Tenove C., Buffie J., McKay S. et Moscrop D., *Digital threats to democratic elections: how foreign actors use digital techniques to undermine democracy*, janvier 2018, passim, https://democracy2017.sites.olt.ubc.ca/files/2018/01/DigitalThreats_Report-FINAL.pdf.

¹⁶³ Lorsque la Federal Communications Commission (Commission fédérale des communications – FCC) a envisagé d'abroger certaines règles régissant internet en 2017, 21 millions des commentaires sur les 22 millions reçus par la Commission sur son site web étaient de fausses nouvelles (Manheim K. & Kaplan L., *op. cit.*, p. 145.)

¹⁶⁴ Sunstein C. R., *op. cit.*

¹⁶⁵ Mor N., « No Longer Private: On Human Rights and the Public Facet of Social Network Sites », *Hofstra Law Review* 47 (2018), p. 669, <https://www.hofstralawreview.org/wp-content/uploads/2019/04/bb.7.mor.pdf> (6 août 2020).

¹⁶⁶ *Ibidem*, p. 42 ff.

leur propre profil, mais ils ne pouvaient pas les « bloquer », car cela aurait empêché certains citoyens d'être informés sur des questions d'intérêt général¹⁶⁷.

2.6. Les médias, chiens de garde de la surveillance ?

L'analyse des données massives a contribué au développement des techniques de reconnaissance faciale artificielle. Grâce aux capacités de l'IA, les logiciels peuvent parcourir et comparer une énorme quantité d'images, pour trouver des « correspondances » (*matches*). Contrairement aux anciennes caméras en « circuit fermé, que les agents humains examinent à la recherche de correspondances, la vision par ordinateur a aujourd'hui la capacité de traiter les images presque instantanément. Dans une décision de 2019, un tribunal gallois s'est penché sur la reconnaissance faciale artificielle¹⁶⁸. Le logiciel que la police galloise avait déployé lors de plusieurs événements publics était capable de traiter jusqu'à 40 visages par seconde. Le chiffre total est impressionnant : en une cinquantaine de déploiements, le logiciel a traité environ 500 000 personnes, soit une personne sur six de la population totale du pays de Galles. L'IA peut devenir un puissant outil de surveillance de masse, comme cela s'est déjà produit dans des pays comme la Chine, où un projet de système de surveillance systématique basé sur l'IA, avec plus d'un demi-milliard de caméras déployées, est en cours¹⁶⁹.

La reconnaissance faciale recoupe une série de questions abordées ci-dessus. Premièrement, les techniques de reconnaissance faciale relèvent de la vie privée. Elles traitent les visages humains – pas seulement ceux qui figurent dans une base de données, mais ceux de tout le monde. En fait, pour exclure une personne du groupe de personnes d'intérêt, un logiciel doit d'abord traiter son visage. Selon la culture juridique européenne, une intrusion aussi massive dans la vie privée doit être dûment justifiée. Comme la Cour européenne des droits de l'homme l'a souligné à plusieurs reprises, les intérêts publics ne priment pas sur les préoccupations relatives à la vie privée – au contraire, ils exigent une évaluation préliminaire des avantages et des coûts attendus pour garantir que tout déploiement est proportionné à l'objectif poursuivi¹⁷⁰.

Deuxièmement, les techniques de reconnaissance faciales risquent d'être biaisées. Comme indiqué ci-dessus, les « faux positifs » – les mauvaises correspondances – sont plus fréquents dans les groupes ethniques qui sont sous-représentés dans le matériel de

¹⁶⁷ *Knight First Amendment Inst. at Columbia Univ. v. Trump* 302 F. Supp. 3d 541 (SDNY 2018), <https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=2780&context=historical> (6 août 2020).

¹⁶⁸ (*Bridges*) v. *The Chief Constable of South Wales Police et al.*, [2019] EWHC 2341, <https://www.judiciary.uk/wp-content/uploads/2019/09/bridges-swp-judgment-Final03-09-19-1.pdf>.

¹⁶⁹ Carter W. M., « Big Brother facial recognition needs ethical regulations », *The Conversation*, 22 juillet 2018, <https://theconversation.com/big-brother-facial-recognition-needs-ethical-regulations-99983>.

¹⁷⁰ *Lopez Ribalda et autres c. Espagne* (requêtes n° 1874/13 et 8567/13 : <http://hudoc.echr.coe.int/fre?i=001-197098>) ; *Gorlov et autres c. Russie* (requête n° 27057/06 ; 56443/09 ; 25147/14 : <http://hudoc.echr.coe.int/spa?i=001-194247>) ; *Antovic et Mirkovic c. Monténégro* (requête n° 70838/13 : <http://hudoc.echr.coe.int/fre?i=001-178904>) ; *Bărbulescu c. Roumanie* (requête n° 61496/08 : <http://hudoc.echr.coe.int/spa?i=001-177082>).

formation¹⁷¹. Les faux positifs ont souvent des conséquences pratiques, car ils peuvent renforcer les préjugés raciaux et orienter les institutions publiques, telles que les patrouilles de police, à cibler les minorités ethniques pour lesquelles le logiciel renvoie davantage de faux positifs¹⁷².

Troisièmement, la reconnaissance faciale peut être trompeuse pour diverses raisons. Certains logiciels sont capables d'exploiter les immenses capacités de l'IA en utilisant des images en direct et enregistrées provenant de n'importe quelle source internet¹⁷³. Cette technologie peut exploiter l'industrie des médias pour rassembler plus de matériel et augmenter sa base de données. Un débat est en cours sur les avantages et les inconvénients liés au développement ou à l'adoption d'un logiciel qui passe au crible le web pour trouver des correspondances, comme cela s'est produit dans de nombreux services de police locaux aux États-Unis pour retrouver des suspects. Un ensemble de données aussi énorme s'appuie sur une multiplicité de matériels qui peuvent être fallacieux, incorporer des biais¹⁷⁴ et enregistrer dans un dossier tout élément de la vie sociale ou médiatique.

Il n'est pas surprenant qu'IBM,¹⁷⁵ Microsoft¹⁷⁶ et Amazon¹⁷⁷ aient récemment déclaré qu'ils ne proposeraient plus leurs technologies de reconnaissance faciale à la police. De nombreux États américains envisagent d'interdire la reconnaissance faciale artificielle ou ont déjà mis en œuvre une législation qui la limite ou l'interdit¹⁷⁸. Il existe donc un consensus croissant dans les pays occidentaux sur le fait que même les intérêts publics ne peuvent justifier des systèmes de surveillance de masse omniprésents qui exploitent le web.

¹⁷¹ Buolamwini J. & Gebru T., « Gender shades: Intersectional accuracy disparities in commercial gender classification » *Proceedings of Machine Learning Research* 81, 2018, pp. 1 et 15, <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.

¹⁷² Fung B. et Metz R., « This may be America's first known wrongful arrest involving facial recognition », 24 juin 2020, *CNN Business*, <https://edition.cnn.com/2020/06/24/tech/aclu-mistaken-facial-recognition/index.html>.

¹⁷³ Hill K., « The secretive company that might end privacy as we know it », *New York Times*, 18 janvier 2020, <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> ; Ducklin P., « Clearview AI facial recognition sued again – this time by ACLU », *Naked Security*, 29 May 2020, <https://nakedsecurity.sophos.com/2020/05/29/clearview-ai-facial-recognition-sued-again-this-time-by-aclu>.

¹⁷⁴ Geiger R. S. et al., « Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from? », <https://arxiv.org/abs/1912.08320>.

¹⁷⁵ Krishna A., « IBM CEO's Letter to Congress on Racial Justice Reform », 8 juin 2020, <https://www.ibm.com/blogs/policy/facial-recognition-susset-racial-justice-reforms/>.

¹⁷⁶ Greene J. « Microsoft won't sell police its technology, following similar moves by Amazon and IBM », *The Washington Post*, 11 juin 2020, <https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/>.

¹⁷⁷ Hao K., « The two-year fight to stop Amazon from selling face recognition to the police », *MIT Technology Review*, 12 juin 2020, <https://www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight>. Voir également Hartzog W., *op. cit.*, pp. 76-77.

¹⁷⁸ Voir la loi de l'Illinois relative au caractère privé des informations biométriques, <https://www.termsfeed.com/blog/bipa/>.

2.7. Le marché des médias : des stratégies de marché dirigées par les données massives

Les données massives ont révolutionné l'univers des médias. De nombreux acteurs de l'industrie des médias dépendent désormais des grandes entreprises technologiques pour mieux se connecter à leur public¹⁷⁹. En fait, la collecte et le traitement d'énormes quantités de données de manière fructueuse nécessitent des capacités que peu de sociétés possèdent. Le nombre d'entreprises capables de récolter des données massives est très limité, et la majorité des acteurs du marché s'appuient sur ce pool pour mieux comprendre qui sont leurs clients, quel type de stratégie de marché ils doivent mettre en œuvre ou comment gagner en visibilité. Certaines grandes entreprises technologiques de ce domaine, comme Amazon, produisent même des contenus médiatiques elles-mêmes. Grâce à leurs capacités technologiques, les grandes entreprises technologiques opèrent donc désormais comme des fabricants de médias et/ou comme des médiateurs entre l'industrie des médias et ses consommateurs.

L'affaire Google Spain¹⁸⁰ jugée par la Cour de justice de l'Union européenne, qui a fait date, résume le rôle central que jouent désormais les grandes entreprises technologiques dans le domaine de l'information et leur résistance aux lois qui le régissent. Lorsqu'un individu s'est plaint qu'une recherche Google sur son nom renvoyait une liste de résultats en haut de laquelle figurait un très ancien article de journal le concernant qui pouvait encore ruiner sa réputation, la première ligne de défense de Google a été qu'elle ne traitait pas de données à caractère personnel ; elle ne faisait qu'associer des recherches à des résultats¹⁸¹. En d'autres termes, Google a fait valoir qu'elle n'était pas responsable des informations qu'elle rendait disponible par le biais d'une recherche Google. La Cour a rendu un arrêt historique, attestant le rôle unique joué par Google dans les recherches sur internet. Elle a estimé que Google était responsable de la manière dont elle classait ses réponses à une requête, car cela pouvait ressusciter des informations oubliées depuis longtemps et qui n'auraient pas été accessibles au grand public autrement.

Les grandes entreprises technologiques ne se contentent pas de peupler le marché des médias. Elles affectent aussi profondément sa dynamique. Leur capacité unique à profiler le marché piège leurs utilisateurs dans un phénomène de « blocage » et génère un quasi-monopole de marché¹⁸². Elles sont si omniprésentes et indispensables que ceux qui ne veulent pas les utiliser doivent souvent quitter le marché. De nombreux internautes savent que « la visite d'un seul site web entraîne généralement la divulgation du comportement de navigation à plus de 100 tiers qui cherchent à limiter leur propre responsabilité juridique au moyen de "politiques de confidentialité" denses qui peuvent s'étendre sur des centaines de pages », mais ils ne peuvent éviter de visiter les mêmes sites

¹⁷⁹ Tsesis T., *op. cit.*, p. 1589.

¹⁸⁰ *Google Spain SL et al. c. Agencia Española de Protección de Datos*, C-131/12, <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:62012CJ0131&from=FR>.

¹⁸¹ *Ibid.*, paragraphe 22.

¹⁸² AGCM, AGCOM, et Garante per la protezione dei dati personali, *Indagine conoscitiva sui Big Data*, *op. cit.*, pp. 26 et 78.

web à maintes reprises¹⁸³. Les quelques entreprises qui exploitent le potentiel des données massives peuvent patrouiller encore plus loin sur leur territoire en se livrant à des « acquisitions tueuses », par lesquelles elles achètent des start-ups innovantes soit pour exploiter les données qu'elles ont collectées¹⁸⁴, soit pour protéger leur position dominante¹⁸⁵. Selon les mots de Frank Pasquale, comme « le pharaon essaie de tuer le bébé Moïse », les grandes entreprises technologiques peuvent refuser à leurs rivaux « la possibilité de se développer »¹⁸⁶.

La présence simultanée de plusieurs entreprises qui utilisent des données massives ne garantit pas qu'un marché soit concurrentiel¹⁸⁷. Les données massives peuvent contribuer à l'élaboration de stratégies de marché, notamment en matière de prix, qui profitent aux concurrents et non aux clients. Il est prouvé que les algorithmes de différentes entreprises peuvent maximiser les prix par une stratégie collusoire implicite, simplement en traitant les informations sur le marché lui-même¹⁸⁸. Un algorithme peut suggérer à une entreprise d'augmenter ses prix parce qu'il prédit que ses concurrents décideront de faire de même. Grâce au profilage et au regroupement des utilisateurs, ils peuvent également « segmenter... le marché » et faire payer chaque utilisateur en fonction de ce qu'il est disposé à payer. Ces pratiques créent « un revenu maximum [pour les entreprises] mais pas de bien-être pour les consommateurs »¹⁸⁹. Une telle stratégie de marché dirigée par les données n'est généralement pas sanctionnable, car il n'y a pas de collusion, mais elle présente les avantages qui découlent habituellement des comportements collusoires¹⁹⁰.

2.8. Approches réglementaires des systèmes basés sur l'IA

Nombreux sont ceux qui ont exprimé la nécessité de mettre en place de nouveaux dispositifs réglementaires afin de garantir que l'IA soit utilisée dans le respect de l'État de droit, des droits fondamentaux et des valeurs éthiques. Les grandes entreprises technologiques ont longtemps résisté aux efforts publics visant à réglementer le

¹⁸³ Contrôleur européen de la protection des données, Avis 3/2018 Avis du CEPD sur la manipulation en ligne et les données à caractère personnel, *op. cit.*, p. 7.

¹⁸⁴ Zuboff S., *The Age of Surveillance Capitalism*, Profile Books, 2019, pp. 102-103.

¹⁸⁵ AGCM, AGCOM, et Garante per la protezione dei dati personali, « Indagine conoscitiva sui Big Data », *op. cit.*, p. 81, Voir également Hughes C., *op. cit.*

¹⁸⁶ Pasquale F., *The Black Box Society*, Harvard University Press, 2015, p. 67.

¹⁸⁷ Contrôleur européen de la protection des données, Opinion 3/2020 on the European strategy for data, *op. cit.* p. 8 (où il est mis en garde contre la création ou le renforcement de « situations d'oligopole de données »).

¹⁸⁸ Den Boer A. V., « Dynamic pricing and learning: Historical origins, current research, and new directions », *Surveys in operations research and management science*, 20, 2015, p. 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2334429 ; AGCM, AGCOM, et Garante per la protezione dei dati personali, « Indagine conoscitiva sui Big Data », *op. cit.*

¹⁸⁹ Délégué européen à la protection des données, Avis 8/2016 Avis du CEPD sur une application cohérente des droits fondamentaux à l'ère des données massives (*Big Data*), *op. cit.* p. 6.

¹⁹⁰ Harrington, J. E. Jr., « Developing competition law for collusion by autonomous artificial agents », *Journal of Competition Law & Economics*, 14, 2019, pp. 349-351, <https://academic.oup.com/jcle/article-abstract/14/3/331/5292366?redirectedFrom=fulltext>.

domaine¹⁹¹, mais semblent maintenant avoir accepté la nécessité de limiter l'IA, bien qu'elles privilégient l'autorégulation des entreprises aux règles imposées par les États.

Toutefois, la plupart des contraintes ne visent pas à freiner l'utilisation de l'IA ; en fait, elles sont censées renforcer son rôle en la rendant plus fiable et digne de confiance¹⁹². Il existe en effet un large consensus sur le fait que l'IA doit être « légale » (conforme à la loi), « éthique » (s'engager à respecter les principes et valeurs éthiques) et « robuste » (technologiquement et sociologiquement sûre), afin de s'intégrer avec succès dans les sociétés humaines¹⁹³.

Les débats soulignent souvent que les analyses de données massives ont besoin d'une nouvelle approche en matière de réglementation juridique. Les outils traditionnels peuvent ne pas suffire à garantir que le monde des données massives respecte les valeurs humaines fondamentales. En raison de la structure de boîte noire de l'IA et de ses effets à grande échelle, les sanctions juridiques ne sont guère capables de contraindre les technologies et stratégies basées sur les données massives. Les poursuites judiciaires peuvent arriver tardivement, lorsque la réputation d'une personne ou d'une entreprise est en ruine, et les responsabilités peuvent être difficiles à localiser. L'IA doit intégrer des valeurs juridiques dans son traitement des données, afin de s'assurer qu'elle les protège pendant son fonctionnement.

En raison de la richesse des informations qu'elle recueille, de son déploiement omniprésent et de sa capacité à remplacer les opérateurs humains par des robots, l'IA pose également des questions éthiques. L'*éthique numérique* est une nouvelle frontière pour la réglementation de l'IA et attire fortement l'attention, en particulier aux États-Unis, au Canada et en Europe, où les codes éthiques se multiplient¹⁹⁴. En tant que domaine, l'éthique numérique couvre une multitude de sujets, notamment « les problèmes moraux liés aux données et aux informations », les algorithmes ... et les pratiques et infrastructures correspondantes »¹⁹⁵, d'une manière qui recoupe différentes disciplines et perspectives. Bien que le débat soit extrêmement animé, la situation évolue peu en ce moment, notamment en raison des difficultés à différencier les composantes juridiques et éthiques de la réglementation de l'IA¹⁹⁶.

¹⁹¹ Zuboff S., *op. cit.* p. 105.

¹⁹² Van Dijk N. & Casiraghi S., « The ethicification of privacy and data protection law in the European Union: The case of artificial intelligence », *Brussels Privacy Hub*, 6, 22, mai 2020, p. 5, <https://brusselsprivacyhub.eu/publications/BPH-Working-Paper-VOL6-N22.pdf>.

¹⁹³ Groupe d'experts de haut niveau sur l'intelligence artificielle, Lignes directrices en matière d'éthique pour une IA digne de confiance, p. 2, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>. Voir N. van Dijk & S. Casiraghi, *op. cit.*, p. 14.

¹⁹⁴ Jobin A., Ienca M. et Vayena E., « The global landscape of AI ethics guidelines », *Nature Machine Intelligence*, 1, 2019, pp. 393-395, <https://www.nature.com/articles/s42256-019-0088-2>.

¹⁹⁵ Floridi L., *op. cit.*, p. 3.

¹⁹⁶ Par exemple, voir la Recommandation CM/Rec(2020)1 du Comité des Ministres aux États membres sur les impacts des systèmes algorithmiques sur les droits de l'homme du Conseil de l'Europe, 8 avril 2020, https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154, qui présente la variété des couches réglementaires nécessaires au développement de systèmes fondés sur l'IA.

2.9. Conclusion

Les données massives sont une des raisons principales du succès sociétal, économique et politique de l'IA. Le traitement d'énormes quantités de données est crucial pour les grandes entreprises technologiques. Cependant, tout n'est pas parfait, et les personnes impliquées dans le secteur doivent prendre des mesures pour que l'IA soit bénéfique pour l'être humain¹⁹⁷. Chris Hughes, co-fondateur de Facebook, a averti que la numérisation de l'économie pourrait contribuer à ce qu'il perçoit comme « un déclin de l'esprit d'entreprise, une stagnation de la croissance de la productivité, des prix plus élevés et moins de choix pour les consommateurs »¹⁹⁸. Les enjeux sont tels qu'un député français a même soumis une proposition visant à inscrire une *Charte de l'intelligence artificielle et des algorithmes* dans le préambule de la constitution française, afin de mieux protéger les droits de l'homme¹⁹⁹.

L'IA maximise la mobilisation des individus. Obtenir « autant de réponses que possible du plus de personnes possibles » est un facteur clé de succès, car cela permet de fournir un retour d'information et donne aux entreprises la possibilité d'adapter en temps réel leurs plans et modèles commerciaux à leurs clients²⁰⁰. Les acteurs politiques et influenceurs sociaux exploitent ce phénomène en déclenchant des réactions émotionnelles chez leur public potentiel. La politique et l'économie des données massives placent les médias au centre de la scène, car ils diffusent les nouvelles, recueillent les informations, traitent les émotions et relient les sphères sociales.

Les données massives amplifient le rôle des médias pour les sociétés contemporaines. Les entreprises, les responsables politiques, les décideurs et autres personnalités politiques exploitent les données massives pour promouvoir leurs idées, leurs programmes et leurs opinions, ainsi que pour influencer leur auditoire²⁰¹. Les plateformes internet permettent aux médias traditionnels de diffuser leur contenu et de générer une nouvelle concurrence entre les médias traditionnels et les nouveaux médias.

Les acteurs médiatiques peuvent également jouer un rôle négatif. En profilant les « schémas de pensée et la composition psychologique », ils peuvent délibérément désinformer un public et l'induire en erreur²⁰². En outre, dans les pays où peu d'acteurs médiatiques opèrent, ou dans lesquels il n'existe que des médias sociaux gérés par l'État ou presque²⁰³, un régime politique peut contrôler efficacement les informations et la façon dont les gens y réagissent, en diffusant des réactions favorables fabriquées de toutes pièces

¹⁹⁷ Voir les Principes d'Asilomar, développés en conjonction avec la conférence d'Asilomar de 2017. *Future of Life Institute*, <https://futureoflife.org/ai-principles>.

¹⁹⁸ Voir également Hughes C., *op. cit.*

¹⁹⁹ http://www.assemblee-nationale.fr/dyn/15/textes/15b2585_proposition-loi.

²⁰⁰ Akin Unver H., « Artificial intelligence, authoritarianism and the future of political systems », Centre for Economics and Foreign Policy Studies, juillet 2019, p. 3, https://edam.org.tr/wp-content/uploads/2018/07/AKIN-Artificial-Intelligence_Bosch-3.pdf.

²⁰¹ *Idem*.

²⁰² Contrôleur européen de la protection des données, Avis 4/2015. Vers une nouvelle éthique numérique, *op. cit.*, p. 7.

²⁰³ Pasquale F., *op. cit.* p. 10, note que « la distinction entre l'État et le marché s'estompe » en raison du déploiement massif de l'IA dans des secteurs stratégiques d'intérêt public et privé.

et en isolant les commentaires défavorables²⁰⁴. Dans le scénario généré par les données massives, les médias peuvent jouer un rôle majeur en protégeant la démocratie, l'égalité, les groupes minoritaires et les sociétés ouvertes – ou en les affaiblissant²⁰⁵.

Enfin, la surveillance de masse peut avoir un effet dissuasif sur la créativité et l'innovation. Malgré les attentes antérieures selon lesquelles l'IA ne ferait que stimuler l'inventivité²⁰⁶, certains observent « une tendance à décourager ou à pénaliser la spontanéité, l'expérimentation ou la déviation par rapport à la “norme” statistique et à récompenser un comportement conformiste »²⁰⁷.

Le vaste déploiement de l'IA auquel nous assistons exige que la sphère médiatique prenne conscience du rôle unique qui est le sien. Le secteur des médias doit s'efforcer d'utiliser l'IA de manière légale, éthique et robuste. Grâce à leur rôle de liaison, les médias pourraient encourager le monde plus large des entreprises basées sur l'IA à adopter les mêmes valeurs et à devenir légales, éthiques et robustes. En particulier, un engagement éthique peut encourager les plateformes médiatiques à ne pas se limiter à un rôle purement passif. Bien que de nombreuses réglementations limitent la responsabilité juridique des fournisseurs vis-à-vis du contenu qu'ils hébergent²⁰⁸, et que les charges supplémentaires imposées aux médias n'aient pas réussi à encourager un renforcement du contrôle, les plateformes médiatiques peuvent néanmoins se fixer l'objectif éthique louable de surveiller leur contenu²⁰⁹.

²⁰⁴ Akin Unver H., *op. cit.*, p. 8. Voir également Meaker M., « How governments use the Internet to crush online dissent », *The Correspondent*, 27 novembre 2019, <https://thecorrespondent.com/142/how-governments-use-the-internet-to-crush-online-dissent/18607103196-db0c0dab>.

²⁰⁵ Groupe d'experts de haut niveau sur l'intelligence artificielle, « Lignes directrices en matière d'éthique pour une IA digne de confiance », *op. cit.*, p. 11.

²⁰⁶ Perritt, H. H., Jr., *op. cit.*, p. 107.

²⁰⁷ Contrôleur européen de la protection des données, Avis 4/2015. Vers une nouvelle éthique numérique, *op. cit.*, p. 9. Voir également Pan S. B., *op. cit.*, p. 257 (« The goal of big data is to generalize ») et Pasquale F., *op. cit.*, p. 188.

²⁰⁸ Perritt H. H., Jr., *op. cit.*, p. 149.

²⁰⁹ ERGA2020 Subgroup 1 – Enforcement, ERGA Position Paper on the Digital Services Act, p. 6, https://nellyo.files.wordpress.com/2020/06/erga_sq1_dsa_position-paper_adopted-1.pdf.

La liberté d'expression, la diversité et le pluralisme

*L'usage de l'IA fait naître des préoccupations d'une importance particulière pour le domaine des médias, relatives à la diversité et au pluralisme. Et pourtant, pour ceux qui sont en âge de se souvenir du temps où le nombre de chaînes de TV d'un pays donné se comptait sur les doigts d'une main et où les journaux se lisaient sur du papier journal, ce problème de diversité (du moins du point de vue quantitatif) peut sembler quelque peu exagéré. Il existe de nos jours quantité de chaînes de TV et tous les journaux du monde sont à la portée d'un simple clic. On pourrait même dire que le seul obstacle qui empêche aujourd'hui quelqu'un d'accéder à l'ensemble de l'information mondiale n'est pas le recours aux algorithmes, mais plutôt l'accès payant de certains services. Mais c'est précisément parce que l'offre d'information est extraordinairement large que les gens ont besoin de filtres. Et comme nous l'avons souligné plus haut, l'IA ne manque pas d'efficacité lorsqu'il s'agit de filtrage. Les services de vidéo à la demande ou d'information peuvent personnaliser l'information proposée à chaque utilisateur d'internet, en fonction des programmes qu'il a regardés par le passé, des articles qu'il a lus, ou d'autres préférences. Cela ne va pas sans poser de problèmes : les bulles dites de filtres qui surviennent lorsque les algorithmes filtrent « les faits et les différents points de vue, renforçant ainsi les points de vue profondément ancrés, voire les préjugés. »²¹⁰ Néanmoins, tout le monde ne s'accorde pas sur l'existence et les effets de ces bulles de filtres. Dans sa contribution à cette publication, **Mira Burri** admet que les médias sur mesure ont certaines répercussions sur la diversité et qu'il est nécessaire d'être attentifs au pouvoir des plateformes, mais elle souligne également ses doutes quant à leur impact direct sur la fragmentation du discours public et la polarisation des opinions²¹¹. Même les partisans de la théorie de la bulle de filtres admettent ne pas pouvoir démontrer son existence réelle²¹² et que les preuves concrètes de l'existence de ces bulles sont pour l'instant plutôt limitées²¹³. **Sarah Eskens**, dans sa contribution à cette publication, affirme : « [L]e défi qui se pose actuellement pour les médias d'information et les autorités publiques est l'élaboration de codes de déontologie journalistique, de normes d'autorégulation, et peut-être même d'une réglementation propre à maîtriser les risques que*

²¹⁰ See Andrea Pin's contribution to this publication. Voir la contribution d'Andrea Pin à cette publication.

²¹¹ Pour d'autres points de vue critiques sur ce sujet, voir par exemple Bruns A., « It's Not the Technology, Stupid : How the 'Echo Chamber' and 'Filter Bubble' Metaphors Have Failed Us », <http://snurb.info/node/2526>.

²¹² Zuiderveen Borgesius, F., Trilling, D., Moeller, J., Bodó, B., de Vreese, C. H., & Helberger, N., « Should we worry about filter bubbles ? » *Internet Policy Review*, 5(1), <https://doi.org/10.14763/2016.1.401>.

²¹³ Helberger N., Eskens S., van Drunen M., Bastian M., Moeller J., Les incidences des applications d'IA dans les médias sur la liberté d'expression, <https://rm.coe.int/prems-044020-fra-2018-ai-and-freedom-of-expression-txt-a4-web/16809e290e>.



soulève l'IA pour la liberté d'expression, tout en lui permettant de contribuer au débat public, au pluralisme des médias, à la libre circulation de l'information et à d'autres objectifs sociétaux. »

3. Les conséquences du recours à l'intelligence artificielle par les médias d'information sur la liberté d'expression

Sarah Eskens, Université d'Amsterdam

3.1. Introduction

Les médias d'information s'appuient de plus en plus sur l'intelligence artificielle (IA) dans le cadre de leurs activités. En 2018, l'Institut Reuters pour l'étude du journalisme a interrogé près de 200 acteurs importants du journalisme. Près des trois quarts d'entre eux ont indiqué qu'ils avaient déjà recours à l'IA dans leur organisation²¹⁴.

L'utilisation de l'IA présente des opportunités pour les médias d'information et peut les aider à remplir leur rôle démocratique. Un rapport de l'Union européenne de radio-télévision a aussi souligné les opportunités qu'offre l'IA pour le journalisme de service public²¹⁵. Le recours à l'IA par les médias d'information pourrait ainsi tomber dans le champ de la protection de la liberté d'expression des médias, ce qui est un facteur important compte tenu des appels à réglementer l'IA dans plusieurs domaines. Dans le même temps, l'utilisation de l'IA par les médias d'information pourrait avoir un impact sur la capacité des autres participants au débat public à exercer leur droit à la liberté d'expression. Par exemple, les organismes d'information peuvent s'appuyer sur des algorithmes pour modérer les commentaires sur leurs sites web. Si la modération automatisée présentait un biais en faveur, par exemple, de l'anglo-américain standard, elle pourrait empêcher certaines voix de se faire entendre dans le débat public.

Alors que les législateurs débattent de la nécessité d'encadrer l'IA, il convient de s'interroger sur la mesure dans laquelle l'usage de l'IA par les médias d'information peut être réglementé et sur l'équilibre à trouver entre la liberté des médias d'une part et les autres droits et intérêts liés au recours à l'IA par les médias d'information d'autre part. La façon dont les médias d'information utilisent l'IA façonne notre environnement informationnel et pourrait avoir des effets importants sur le débat public, le pluralisme et la diversité de l'information, la libre circulation de l'information et d'autres valeurs

²¹⁴ Newman N., *Journalism, media, and technology trends and predictions 2018*, Institut Reuters pour l'étude du journalisme, p. 29,

<https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-01/RISJ%20Trends%20and%20Predictions%202018%20NN.pdf>.

²¹⁵ Union européenne de radio-télévision, « The next newsroom: Unlocking the power of AI for public service journalism », <https://www.ebu.ch/publications/news-report-2019>.

publiques attachées aux médias d'information²¹⁶. Les questions que soulève l'utilisation de l'IA par les médias d'information sont uniques en raison du rôle démocratique qui leur incombe. Ce chapitre se concentre donc sur l'usage de l'IA par les médias d'information, et non par les autres médias, comme les médias de divertissement, qui exige de trouver un équilibre des intérêts différent²¹⁷.

Ce chapitre est structuré de la manière suivante : il présente le cadre général de la protection de la liberté d'expression ; il examine la mesure dans laquelle le recours à l'IA relève du champ de la liberté des médias, s'interroge sur le type de nouveaux acteurs qui recourent à l'IA et peuvent se prévaloir de la liberté des médias, et sur les devoirs et obligations qui pèsent sur les médias d'information lorsqu'ils utilisent l'IA. Il décrit ensuite comment certaines applications d'IA utilisées par les médias d'information peuvent limiter les droits à la liberté d'expression d'autres participants au débat public, notamment les utilisateurs d'information et les citoyens ou personnalités politiques qui font entendre leur voix grâce aux médias. L'objectif n'est pas de présenter ici un inventaire complet des risques posés par l'utilisation de l'IA par les médias d'information. Il s'agit plutôt d'illustrer les défis que pose l'IA pour la liberté d'expression afin de discuter en substance des droits fondamentaux des différents participants au débat public. Enfin, ce chapitre examine les obligations, en matière de liberté d'expression, qui s'imposent aux États au regard de l'utilisation de l'IA par les médias d'information. Mais avant d'aborder ces questions, il expose brièvement les principaux objectifs des médias d'information lorsqu'ils utilisent l'IA.

3.2. Les applications de l'IA au service des médias d'information

Dans le cadre de ce chapitre, comme d'autres études ayant porté sur la question de l'IA et des médias d'information, nous définissons l'IA au sens large comme « un ensemble d'idées, de technologies et de techniques qui concernent la capacité d'un système informatique à réaliser des tâches qui requièrent normalement une intelligence humaine²¹⁸ ». L'IA est donc un terme générique qui englobe diverses technologies numériques, comprenant entre autres l'apprentissage automatique, la reconnaissance des images, le traitement automatique des langues, et la génération automatique de langage naturel. Les autres

²¹⁶ Conseil de l'Europe, Recommandation CM/Rec(2018)1 du Comité des Ministres aux États membres sur le pluralisme des médias et la transparence de leur propriété, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680790e36 ; Conseil de l'Europe, Recommandation CM/Rec(2015)6 du Comité des Ministres aux États membres sur la libre circulation transfrontière des informations sur internet, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805c3f76.

²¹⁷ Voir chapitre 4 de cette publication.

²¹⁸ Beckett C., « New powers, new responsibilities: A global survey of journalism and artificial intelligence », London School of Economics and Political Science, p. 16, <https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/>.

chapitres de cette publication illustrent la diversité des applications pour les médias qui sont considérées comme relevant de l'IA.

Dans ce chapitre, on distingue quatre objectifs pour lesquels les médias d'information peuvent utiliser l'IA : la collecte d'informations, la production d'informations, la distribution d'informations, et la modération des commentaires des utilisateurs²¹⁹. Les trois premiers objectifs relèvent d'un processus journalistique classique, et le quatrième objectif résulte de la possibilité qu'offrent parfois les médias d'information de poster des commentaires sur leurs sites web. Comme souligné dans un récent rapport de l'Assemblée parlementaire, la plupart des discussions sur l'utilisation de l'IA dans les processus de communication en ligne ont porté sur la modération du contenu, tandis que la manière dont l'IA façonne l'environnement de l'information en ligne est tout aussi importante²²⁰.

Au point de départ du processus journalistique, les médias d'information peuvent recourir à l'IA pour la collecte d'informations. L'IA peut en effet être utilisée pour identifier des informations et événements importants, générer des idées de sujets, et suivre des événements et des questions d'actualité. Une fois que les journalistes ont recueilli des informations sur des sujets éventuels, ils peuvent utiliser l'IA pour la production d'informations. Il s'agit par exemple de l'usage de l'IA pour la rédaction de sujets (parfois appelée « journalisme automatisé »)²²¹, la création d'images et de vidéos, la vérification des informations, ou la réadaptation de contenus pour de nouveaux publics. Les médias d'information peuvent enfin recourir à l'IA lors de la dernière étape du processus journalistique, à savoir la distribution de l'information. Il s'agit du recours à l'IA pour proposer des recommandations personnalisées, trouver de nouveaux publics, promouvoir l'image de marque du média, et vendre des abonnements. Le recours à la personnalisation a souvent un double objectif. Celle-ci aide les médias d'information à mieux servir leurs utilisateurs, et, notamment pour les médias de service public, à remplir leur mission de service public²²². Elle les aide également à conserver leurs abonnés, à renforcer l'engagement des utilisateurs, et à générer ainsi des revenus de vente et de publicité²²³.

Alors que les médias d'information ont donné la possibilité à leurs lecteurs de poster des commentaires sur leurs plateformes en ligne, l'IA peut leur permettre d'assurer une modération plus efficace de ces commentaires. Par exemple, le *New York Times* a mis en place un système qui repose sur l'apprentissage automatique pour modérer en priorité certains commentaires et en approuver de façon automatisée²²⁴. Le *New York Times* assure la modération de près de 12 000 commentaires par jour et le système automatisé permet

²¹⁹ Beckett C., *Ibid.*, p. 20.

²²⁰ Assemblée Parlementaire, Rapport: La nécessité d'une gouvernance démocratique de l'intelligence artificielle (24 septembre 2020), para. 18-19, <https://pace.coe.int/en/files/28742/html>.

²²¹ Dörr K.N., « Mapping the field of algorithmic journalism », *Digital Journalism* 4(6), pp. 700–722, <https://doi.org/10.1080/21670811.2015.1096748>.

²²² Van den Bulck H. et Moe H., « Public service media, universality and personalisation through algorithms: Mapping strategies and exploring dilemmas », *Media, Culture & Society* 40(6), pp. 875–92, <https://doi.org/10.1177/0163443717734407>.

²²³ Bodó B., « Selling news to audiences: A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media », *Digital Journalism* 7(8), pp. 1054–75, <https://doi.org/10.1080/21670811.2019.1624185>.

²²⁴ Etim B., « The Times sharply increases articles open for comments, using Google's technology », *The New York Times*, <https://www.nytimes.com/2017/06/13/insider/have-a-comment-leave-a-comment.html>.

de prolonger l'ouverture de la section commentaires et d'approuver ceux-ci plus rapidement.

Certaines études sur l'usage de l'IA dans le secteur des médias d'information abordent la question de la modération des commentaires dans le contexte de la distribution de l'information²²⁵. Mais la modération automatisée des contenus s'accompagne de risques particuliers pour les droits à la liberté d'expression des personnes qui commentent les articles et prennent ainsi part au débat public. S'agissant du recours à l'IA pour d'autres aspects de la distribution, par exemple la personnalisation et le marketing, les utilisateurs d'information sont la cible d'une communication mais ils n'en sont pas eux-mêmes des participants actifs. Ce chapitre considère donc l'IA appliquée à la modération de contenus sur les plateformes d'information comme une catégorie distincte. Après cet état des lieux des divers usages de l'IA dans les médias d'information, la prochaine section s'interroge sur la mesure dans laquelle le recours à l'IA est protégé par la liberté des médias.

3.3. Le recours à l'IA par les médias d'information comme élément de la liberté des médias

En Europe, la Convention européenne des droits de l'homme (CEDH) constitue le socle juridique du droit fondamental à la liberté d'expression. Selon l'article 10(1) de la CEDH, toute personne a droit à la liberté d'expression, ce qui comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques.

La Cour européenne des droits de l'homme a été créée afin de garantir que les États respectent leurs obligations découlant de la CEDH. Elle est à l'origine d'une abondante jurisprudence, dans laquelle elle interprète et développe le droit à la liberté d'expression. Dans l'une de ses premières affaires sur la liberté d'expression, la Cour a affirmé que la liberté d'expression constituait l'un des fondements d'une société démocratique et du développement de chacun²²⁶.

3.3.1. Rôle démocratique des médias d'information

Les universitaires spécialistes du journalisme ont mis en lumière plusieurs rôles démocratiques pour les médias d'information²²⁷. En donnant aux citoyens des informations sur la politique et les actualités, les médias sont une source d'information pour le débat démocratique. De plus, ils jouent le rôle de « quatrième pouvoir » en passant au crible l'exercice du pouvoir par les gouvernements, les entreprises et d'autres acteurs influents.

²²⁵ Beckett C., *Ibid.*, p. 28.

²²⁶ Cour européenne des droits de l'homme, *Handyside c. Royaume-Uni* [1976], 5493/72, paragraphe 49, <http://hudoc.echr.coe.int/fre?i=001-62057>.

²²⁷ McNair B., « Journalism and democracy » in T. Hanitzsch et K. Wahl-Jorgensen (éd.) *The Handbook of Journalism Studies*, Routledge, pp. 237–49.

Les médias ont également une fonction de médiation entre citoyens et personnalités politiques, dans la mesure où ils permettent l'existence d'un espace public dans lequel citoyens et politiques peuvent communiquer par le biais de lettres, d'articles d'opinion, de débats en studio et de contributions à des articles.

La Cour européenne des droits de l'homme a réaffirmé ces diverses fonctions démocratiques des médias. Elle a considéré que les médias d'information avaient comme tâches de distribuer l'information et d'assurer un rôle de contrôle ou de « chien de garde » public²²⁸. À cet égard, la Cour a estimé que la liberté d'expression protégeait tant les activités de recherche de l'information que sa publication²²⁹, et non seulement le contenu des informations mais aussi les moyens techniques de transmission ou de captage²³⁰. De plus, la Cour a établi que l'ouverture d'espaces de débat public était l'une des missions des médias d'information²³¹.

Le recours à l'IA par les médias d'information s'inscrit dans leur rôle démocratique tel que protégé par le droit à la liberté d'expression. Ceux-ci peuvent utiliser l'IA pour rassembler des informations sur des sujets d'actualité, par exemple en analysant les données massives (ou big data) au moyen de l'IA. Ils peuvent également s'en servir pour proposer des informations adaptées à chaque citoyen, selon ses intérêts personnels et ses besoins en information. Les médias d'information peuvent aussi s'appuyer sur l'IA pour suivre et surveiller le comportement de grandes entreprises ou la mise en œuvre de politiques publiques. Enfin, ils peuvent également, grâce à l'IA, améliorer le forum de débat public en modérant automatiquement les commentaires de lecteurs.

En raison de leurs fonctions démocratiques, les médias d'information bénéficient d'une protection particulière de leur liberté d'expression. La Cour européenne des droits de l'homme a affirmé que la liberté d'expression revêtait une importance spéciale pour les médias d'information²³². En ce qui les concerne, la Cour parle donc de la « liberté de la presse »²³³ autrement appelée liberté des médias²³⁴. La Cour a estimé que les autorités publiques disposaient d'une marge d'appréciation plus étroite pour déterminer s'il existe un besoin social impérieux d'interférer avec la liberté des médias²³⁵, par rapport à celle dont ils disposent au regard de la liberté d'expression d'autres types d'intervenants. De plus, la Cour a établi que la liberté des médias protégeait les médias d'information contre l'influence de groupes politiques et économiques puissants dans la société, et garantissait

²²⁸ Cour européenne des droits de l'homme, *Barthold c. Allemagne* [1985], 8734/79, paragraphe 58, <http://hudoc.echr.coe.int/eng?i=001-61990>.

²²⁹ Cour européenne des droits de l'homme, *Dammann c. Suisse* [2006], 77551/01, paragraphe 52, <http://hudoc.echr.coe.int/eng?i=001-75174>.

²³⁰ Cour européenne des droits de l'homme, *Autotronic AG c. Suisse* [1990], 12726/87, paragraphe 47, <http://hudoc.echr.coe.int/eng?i=001-62187>.

²³¹ Cour européenne des droits de l'homme, *Társaság a Szabadságjogokért c. Hongrie* [2009], 37374/05, paragraphe 27, <http://hudoc.echr.coe.int/eng?i=001-92179>.

²³² Cour européenne des droits de l'homme, *Sunday Times c. Royaume-Uni (No. 1)* [1979], 6538/74, paragraphe 65, <http://hudoc.echr.coe.int/eng?i=001-62140>.

²³³ Cour européenne des droits de l'homme, *Sunday Times c. Royaume-Uni (No. 1)*, paragraphe 66.

²³⁴ Oster J., *Media Freedom as a Fundamental Right*, Cambridge University Press, p. 48.

²³⁵ Cour européenne des droits de l'homme, *Busuioc c. Moldova* [2004], 61513/00, paragraphe 65, <http://hudoc.echr.coe.int/eng?i=001-67745>.

leur liberté éditoriale²³⁶. Dans le contexte du recueil d'information, la liberté des médias protège les sources des journalistes²³⁷, et les médias peuvent avoir un droit d'accès aux informations dont disposent les autorités publiques²³⁸. Pour autant que l'IA entre dans le champ de la liberté des médias, les autorités publiques sont limitées lorsqu'il s'agit de la réglementer. La prochaine section s'interroge sur les acteurs qui sont susceptibles de se prévaloir de la liberté des médias.

3.3.2. Bénéficiaires de la liberté des médias

De nos jours, l'environnement des médias d'information est formé par un réseau complexe de différents acteurs, notamment les éditeurs de médias d'information, les utilisateurs de médias d'information et les intermédiaires en ligne. Les médias d'information traditionnels et les médias numériques recueillent des informations originales et publient des articles d'actualité par l'intermédiaire de leurs propres organes d'information en ligne ou hors ligne. Avant internet, les éditeurs de médias d'information atteignaient directement leur public dans son ensemble, lorsque les gens achetaient un certain journal ou allumaient une certaine chaîne de télévision ou de radio. Sur internet, les éditeurs de médias d'information peuvent toujours atteindre directement leur public, lorsque les gens consultent leurs sites d'information ou leurs applications. Mais les gens consultent et trouvent de plus en plus d'informations à partir de médias sociaux, de moteurs de recherche ou d'agrégateurs d'information²³⁹. Ces plateformes jouent le rôle d'intermédiaires entre ceux qui éditent l'information et ceux qui la consultent. Tant les médias d'information traditionnels que les médias issus du numérique peuvent manifestement se prévaloir de la liberté des médias. Néanmoins, il convient de s'interroger sur la mesure dans laquelle les intermédiaires en ligne, du fait de leur rôle dans l'environnement informationnel, peuvent également bénéficier de la liberté des médias lorsqu'ils recourent à l'IA.

La Cour européenne des droits de l'homme a établi que divers acteurs pouvaient remplir les fonctions démocratiques incombant habituellement aux médias. Elle a souligné qu'il existait un net intérêt général à autoriser des groupes militants à contribuer au débat public par la diffusion d'informations sur des sujets d'intérêt général²⁴⁰, et a donc examiné le comportement des autorités publiques vis-à-vis de ces groupes à la lumière de la liberté des médias (mais elle a également soutenu que ces groupes devaient respecter certains devoirs et responsabilités habituellement réservés aux médias ; voir prochaine section). Dans une affaire, la Cour a estimé que la création de forums de débat public n'était pas la

²³⁶ Cour européenne des droits de l'homme, *Manole et autres c. Moldova* [2009], 13936/02, paragraphe 98, <http://hudoc.echr.coe.int/fre?i=001-94076>.

²³⁷ Cour européenne des droits de l'homme [GC], *Goodwin c. Royaume-Uni* [1996], 17488/90, paragraphe 39, <http://hudoc.echr.coe.int/eng?i=001-62533>.

²³⁸ Cour européenne des droits de l'homme [GC], *Goodwin c. Royaume-Uni* [1996], 17488/90, paragraphe 39, <http://hudoc.echr.coe.int/eng?i=001-62533>.

²³⁹ Newman N. *et al.*, « Digital News Report 2020 », Institut Reuters pour l'étude du journalisme, pp. 11–12, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.

²⁴⁰ Cour européenne des droits de l'homme, *Steel and Morris c. Royaume-Uni* [2005], 68416/01, paragraphe 89, <http://hudoc.echr.coe.int/eng?i=001-68228>.

seule prérogative des médias d'information professionnels et que des organisations non gouvernementales (ONG) pouvaient aussi jouer un tel rôle²⁴¹. Elle a indiqué que les ONG assumaient par conséquent une fonction de « chien de garde social ». Les activités des organisations de la société civile peuvent donc jouir d'une protection comparable au regard des principes de la CEDH à celle dont bénéficient les médias d'information²⁴². Ces décisions de la Cour européenne des droits de l'homme pourraient constituer une base à partir de laquelle les intermédiaires en ligne, selon le rôle sociétal qu'ils jouent, pourraient être reconnus comme des acteurs comparables aux médias d'information.

Comme nous l'avons vu dans la section précédente, la loi en matière de liberté d'expression reconnaît trois fonctions démocratiques que remplissent les médias d'information : la fourniture d'information au public, la création d'un espace de débat public et leur rôle de « chien de garde ». En s'appuyant sur l'IA, les intermédiaires en ligne peuvent remplir deux de ces fonctions en particulier. Ils peuvent faire en sorte que l'information soit plus facile à trouver et à consulter, grâce à des flux d'information personnalisés et à un accès facile à toute une gamme d'éditeurs de médias d'informations. Les intermédiaires en ligne peuvent aussi créer des forums pour le débat public en permettant aux éditeurs de médias d'information, aux personnalités politiques et aux citoyens de poster des contenus sur leurs plateformes au sein de groupes publics ou privés et en mettant à disposition plusieurs formes d'engagement avec le contenu en ligne, par exemple la fonction « j'aime », les commentaires, ou la possibilité de transférer des articles. Les intermédiaires en ligne peuvent ainsi jouer des rôles proches de ceux des médias d'information et pourraient remplir, tout comme les médias, une fonction de « porte d'entrée » et d'initiateurs du débat public. Le Comité des Ministres du Conseil de l'Europe a également souligné de son côté que les intermédiaires en ligne pouvaient « exercer certaines formes de contrôle qui influencent l'accès des utilisateurs aux informations en ligne, à l'instar des médias », ou encore « assurer d'autres fonctions qui se rapprochent de celles des éditeurs²⁴³ ».

Sur la base de ces principes de liberté d'expression, il est possible d'avancer que les intermédiaires en ligne pourraient se prévaloir de la liberté des médias lorsqu'ils recourent à l'IA pour remplir des fonctions similaires à celles des médias d'information. Tout du moins, les intermédiaires en ligne pourraient bénéficier des droits « normaux » à la liberté d'expression lorsqu'ils facilitent l'accès aux informations par le biais de fils d'information et de résultats de recherches. À cet égard, Van Hoboken évoque la production « d'information sur l'information » par les moteurs de recherche, par exemple lorsque ceux-ci affichent des résultats de recherches à partir du référencement d'informations disponibles par ailleurs. Van Hoboken conclut que la publication de résultats de recherche par un moteur de recherche est protégée par l'article 10 de la CEDH²⁴⁴. De la même manière, on pourrait arguer que la sélection, la hiérarchisation et la personnalisation des fils

²⁴¹ Cour européenne des droits de l'homme, *Társaság a Szabadságjogokért c. Hongrie*, paragraphe 27.

²⁴² Ibid.

²⁴³ Conseil de l'Europe, Recommandation CM/Rec(2018)2 du Comité des Ministres aux États membres sur les rôles et les responsabilités des intermédiaires d'internet, paragraphe 5, https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=0900001680790e37.

²⁴⁴ Van Hoboken J. C. J., « Search engine freedom: On the implications of the right to freedom of expression for the legal governance of Web search engines », Université d'Amsterdam, p. 182, <http://hdl.handle.net/11245/1.392066>.

d'information par les médias sociaux et les agrégateurs d'information au moyen de l'IA constituent bien de « l'information sur l'information » et méritent de bénéficier de la protection de la liberté d'expression.

Dès lors que les médias d'information traditionnels, les médias d'information nés du numérique et enfin les intermédiaires en ligne exercent leur liberté d'expression ou leur liberté des médias, ils sont également tenus de respecter certains devoirs et responsabilités. Ceux-ci font l'objet de la prochaine section.

3.3.3. Devoirs et obligations et codes de déontologie journalistique

Alors que le premier paragraphe de l'article 10 de la CEDH garantit le droit fondamental à la liberté d'expression, le deuxième affirme que l'exercice de la liberté d'expression « comporte certains devoirs et responsabilités » et peut donc être soumis à certaines restrictions prévues par la loi, qui constituent des mesures nécessaires, dans une société démocratique, à un objectif légitime. En d'autres termes, l'article 10 de la CEDH prévoit un mécanisme en vue de garantir que les individus et les organisations exercent leur liberté d'expression de façon responsable. Dans le cadre de ce chapitre, nous nous penchons sur ces devoirs et responsabilités, et non sur les conditions dans lesquelles une interférence avec le droit à la liberté d'expression pourrait être justifiée. La question qui se pose est de définir les devoirs et les responsabilités qui accompagnent l'exercice de la liberté d'expression et ce qu'ils impliquent s'agissant du recours par les médias d'information à l'IA.

Plusieurs acteurs sont soumis à des devoirs et responsabilités lorsqu'ils participent ou contribuent à l'exercice de la liberté d'expression. La Cour européenne des droits de l'homme a affirmé qu'en plus des intervenants ou auteurs eux-mêmes, les personnes et organisations qui fournissent à d'autres auteurs un média ou une plateforme, par exemple les éditeurs²⁴⁵ ou les portails d'information en ligne²⁴⁶, sont aussi tenus à des devoirs et responsabilités quant à la publication et la distribution de contenus issus de tierces parties.

La Cour a estimé que la portée des devoirs et responsabilités imposés à un acteur dépendait de plusieurs facteurs. Tout d'abord, les devoirs et responsabilités d'une personne dépendent de sa situation et des moyens techniques qu'elle utilise pour communiquer²⁴⁷. Dans la mesure où les médias d'information jouent un rôle démocratique singulier, leurs devoirs et responsabilités prennent une importance particulière. La Cour a affirmé que les devoirs et responsabilités des médias d'information revêtaient une importance toute

²⁴⁵ Cour européenne des droits de l'homme, *Éditions Plon c. France* [2004], 58148/00, paragraphe 50, <http://hudoc.echr.coe.int/eng?i=001-66318> ; Cour européenne des droits de l'homme, *Chauvy et autres c. France* [2004], 64915/01, paragraphe 79, <http://hudoc.echr.coe.int/eng?i=001-66418>.

²⁴⁶ Cour européenne des droits de l'homme, *Magyar Tartalomszolgáltatók Egyesülete et Index.hu Zrt c. Hongrie* [2016], 22947/13, paragraphe 62, <http://hudoc.echr.coe.int/eng?i=001-160314> ; Cour européenne des droits de l'homme [GC], *Delfi AS c. Estonie* [2015], 64569/09, paragraphe 113, <http://hudoc.echr.coe.int/eng?i=001-155627>.

²⁴⁷ Cour européenne des droits de l'homme, *Handyside c. Royaume-Uni*, paragraphe 49.

particulière lorsque leurs activités risquent de mettre en péril les droits d'autrui²⁴⁸. En outre, l'impact potentiel du média constitue un critère pour déterminer l'étendue de ses devoirs et responsabilités²⁴⁹. Plus un média d'information a d'impact, plus ses devoirs et ses responsabilités sont lourds. Néanmoins, selon la Cour européenne des droits de l'homme, les devoirs et responsabilités des portails d'information sur internet peuvent dans une certaine mesure se distinguer de ceux des éditeurs traditionnels pour ce qui est des contenus issus de tierces parties²⁵⁰. De manière comparable, le Comité des Ministres du Conseil de l'Europe a recommandé que les devoirs et responsabilités des intermédiaires en ligne soient déterminés, du fait de leurs multiples rôles, au regard des services spécifiques qu'ils fournissent et des fonctions particulières qu'ils exercent²⁵¹.

La Cour européenne des droits de l'homme a estimé que les devoirs et responsabilités des médias d'information impliquaient qu'ils doivent agir « de bonne foi de manière à fournir des informations exactes et dignes de crédit dans le respect de la déontologie journalistique »²⁵². Les codes de déontologie journalistique existaient bien avant que la Cour introduise l'idée que l'exercice de la liberté d'expression s'accompagnait de devoirs et de responsabilités. Dans les années 1920, l'American Society of Newspaper Editors a adopté les « canons du journalisme », qui sont considérés comme l'un des premiers codes déontologiques pour les médias d'information²⁵³. Mais dans les années suivantes, les critiques visant les entreprises de presse ont poussé les gouvernements à envisager une réglementation des médias d'information²⁵⁴. En réponse, ceux-ci ont mis en place des codes déontologiques, des conseils de presse, des médiateurs et d'autres formes d'autorégulation afin d'éviter l'imposition d'une réglementation par le gouvernement²⁵⁵.

En Europe, l'idée que les médias d'information ne devraient pas faire l'objet d'une réglementation concerne principalement la presse écrite. Les États européens réglementent les médias audiovisuels de plusieurs manières, notamment par le biais de la Directive européenne Services de médias audiovisuels. La façon dont la Cour européenne des droits de l'homme a interprété la notion de devoirs et de responsabilités crée désormais également une base juridique normative pour une théorie de la responsabilité sociale de la presse écrite, selon laquelle elle doit se soumettre à une autorégulation et à des codes de déontologie professionnels.

La question est de savoir si les codes déontologiques actuels du journalisme sont à même de traiter la question du recours à l'IA par les médias d'information. En 2001, les universitaires spécialistes du journalisme ont affirmé qu'internet soulevait de nouvelles questions déontologiques pour les journalistes, et conclu que les codes déontologiques

²⁴⁸ Cour européenne des droits de l'homme [GC], *Bladet Tromsø et Stensaas c. Norvège* [1999], 21980/93, paragraphe 65, <http://hudoc.echr.coe.int/eng?i=001-63020>.

²⁴⁹ Cour européenne des droits de l'homme [GC], *Jersild c. Danemark* [1994], 15890/89, paragraphe 31, <http://hudoc.echr.coe.int/eng?i=001-62456>.

²⁵⁰ Cour européenne des droits de l'homme [GC], *Delfi AS c. Estonie*, paragraphe 113.

²⁵¹ Recommandation CM/Rec(2018)2 du Comité des Ministres aux États membres sur les rôles et les responsabilités des intermédiaires d'internet, paragraphe 11.

²⁵² Cour européenne des droits de l'homme [GC], *Bladet Tromsø et Stensaas c. Norvège*, paragraphe 65.

²⁵³ Ward S.J.A., *The invention of journalism ethics: The path to objectivity and beyond*, MQUP, pp. 236–37.

²⁵⁴ Ward S.J.A., *Ibid.*, p. 244.

²⁵⁵ Ward S.J.A., *Ibid.*, p. 245.

traditionnels du journalisme n'étaient pas suffisants pour encadrer la conduite des médias d'information dans l'environnement en ligne²⁵⁶. Plus d'une décennie plus tard, une étude des codes de déontologie journalistique a révélé que la majorité d'entre eux ne comprennent toujours pas de dispositions concernant le journalisme en ligne et les médias numériques²⁵⁷. Helberger et Bastian appellent donc à l'élaboration d'une déontologie du journalisme algorithmique afin de guider les médias d'information lorsqu'ils s'appuient sur l'IA pour la production, la publication et la distribution d'informations²⁵⁸. De la même manière, Dörr et Holnbuchner, deux chercheurs en sciences de la communication, concluent que les organismes de médias devraient adopter des codes déontologiques pour le journalisme algorithmique²⁵⁹.

Il convient de se demander en outre dans quelle mesure les intermédiaires en ligne qui utilisent l'IA pour façonner le cadre de la liberté d'expression en ligne devraient se conformer à des codes déontologiques. Les médias sociaux et les moteurs de recherche tentent depuis longtemps d'échapper à leur responsabilité quant à la manière dont ils sélectionnent et hiérarchisent les informations, en faisant valoir qu'ils ne sont pas des médias. Dans une interview, le PDG de Facebook a souligné que Facebook était « un réseau social » et qu'il préférerait ce terme à celui de « média social », parce que la notion de réseau social met l'accent sur le côté « humain » de la plateforme plutôt que sur son côté « contenus »²⁶⁰. Si les médias sociaux ne sont pas des entités journalistiques, alors ils n'ont pas à respecter les codes de déontologie journalistique.

Il apparaît donc clairement que l'article 10 de la CEDH, tel qu'interprété par la jurisprudence de la Cour, met en place une base juridique normative pour les devoirs et responsabilités des acteurs qui contribuent à la liberté d'expression, qu'ils soient ou non de « véritables » organismes de médias. Dans le même temps, il ressort aussi clairement que les devoirs et responsabilités des intermédiaires en ligne peuvent s'écarter de ceux des médias d'information traditionnels. Cela signifie que les intermédiaires en ligne ne peuvent être contraints de respecter les codes de déontologie journalistique, bien qu'en vertu des principes de la liberté d'expression, il soit clair qu'ils ont une responsabilité lorsqu'ils recourent à l'IA pour réguler le discours sur leurs plateformes et exercer leurs propres droits à la liberté d'expression. Si les intermédiaires en ligne ne mettent pas en place de codes et d'autres instruments d'autorégulation adaptés, alors les gouvernements pourraient à juste titre être amenés à réglementer la manière dont ils exercent leur liberté d'expression par le biais de l'IA. Un raisonnement similaire est apparent dans la manière dont la Commission européenne aborde cette question. Celle-ci menace de réglementer les intermédiaires en ligne s'ils ne se conforment pas à des codes d'autorégulation en matière de désinformation

²⁵⁶ Deuze M. et Yeshua D., « Online journalists face new ethical dilemmas : Lessons from the Netherlands », *Journal of Mass Media Ethics* 16(4), pp. 273–92, https://doi.org/10.1207/S15327728JMM1604_03.

²⁵⁷ Díaz-Campo J. et Segado-Boj F., « Journalism ethics in a digital environment : How journalistic codes of ethics have been adapted to the Internet and ICTs in countries around the world », *Telematics and Informatics* 32(4), pp. 735–44, <https://doi.org/10.1016/j.tele.2015.03.004>.

²⁵⁸ Helberger N. et Bastian M., « AI, algorithms and journalistic ethics », présenté lors de la conférence sur l'avenir du journalisme, Cardiff, 2019.

²⁵⁹ Dörr K. N., « Mapping the field of algorithmic journalism », *Digital Journalism* 4(6), pp. 700–722, <https://doi.org/10.1080/21670811.2015.1096748>.

²⁶⁰ Swisher K., « Zuckerberg : The Recode interview », *Recode*, <https://www.recode.net/2018/7/18/17575156/mark-zuckerberg-interview-facebook-recode-kara-swisher>.

et de fausses informations. Une approche semblable pourrait être suivie s'agissant de la façon dont les intermédiaires en ligne utilisent l'IA pour façonner l'environnement informationnel.

Après cette analyse des principes de la liberté d'expression applicables à l'utilisation de l'IA par les médias d'information et les autres acteurs de l'environnement informationnel en ligne, il s'agit maintenant de se demander de quelle manière le recours à l'IA par les médias d'information affecte les droits à la liberté d'expression des autres participants au débat public.

3.4. Conséquences de l'IA pour les droits à la liberté d'expression des utilisateurs d'information et des autres participants au débat public

L'utilisation de l'IA par les médias d'information à des fins de collecte d'informations, de production d'informations, de distribution d'informations et de modération présente plusieurs risques pour les droits à la liberté d'expression des utilisateurs d'information et des autres intervenants dans l'environnement informationnel médiatique. Il faut noter que le recours à l'IA par les médias d'information est également susceptible de renforcer les droits à la liberté d'expression et le droit de recevoir des informations des utilisateurs d'information et des participants au débat public. Du point de vue de la réglementation, l'enjeu est donc de circonscrire les risques sans priver l'environnement informationnel et le débat public des effets positifs possibles de l'IA.

Au stade de la collecte d'information, les médias d'information peuvent tirer parti de l'IA pour identifier des tendances et des phénomènes au sein des données massives. Utilisée de cette façon, l'IA peut faire ressortir, à partir des données massives, des faits et événements que l'œil humain ne pourrait pas facilement déceler. Toutefois, l'intérêt de l'IA dans le cadre de la recherche d'informations dépend de la disponibilité de données (publiques). Le système de recherche automatisée de sujets est en effet susceptible de passer à côté d'événements ou sujets de société qui ne font pas l'objet d'une large quantité de données²⁶¹. Les voix des individus qui sont impliqués dans des événements et actualités qui ne génèrent pas suffisamment de données massives peuvent être écartées des médias d'information qui reposent sur l'usage de ces données, et ainsi ne pas être entendues dans le débat public.

Au stade de la production d'informations, les médias d'information peuvent utiliser l'IA pour générer du texte et des images, vérifier l'information, traduire automatiquement, rédiger des messages pour les médias sociaux, ou adapter des articles produits pour le plus grand nombre à des publics particuliers. Le recours à l'IA pour la production automatisée d'informations peut générer des contenus illicites qui contreviennent au droit d'autrui, par exemple du discours de haine, des contenus diffamatoires ou qui enfreignent le droit

²⁶¹ Hansen M. *et al.*, « Artificial intelligence : Practice and implications for journalism », Tow Center for Digital Journalism, p. 17, <https://doi.org/10.7916/D8X92PRD>.

d'auteur. La question se pose alors de savoir qui est responsable, y compris du point de vue juridique, de ces contenus illicites : l'organisation de médias qui a décidé de s'appuyer sur l'outil d'IA, le développeur de l'outil d'IA, ou l'IA elle-même ?

Des spécialistes du droit ont avancé qu'il était possible d'affirmer la responsabilité des agents de l'IA²⁶². Des juristes américains ont également avancé que le premier amendement, qui protège le droit constitutionnel à la liberté d'expression, devrait s'appliquer aux discours automatisés²⁶³. Cependant, du point de vue du droit positif, l'IA ne peut avoir de personnalité juridique et donc être tenue pour responsable dans les juridictions européennes. En 2017, le Parlement européen a appelé la Commission européenne à réfléchir à la création d'un statut juridique spécial (« personne électronique ») pour les robots²⁶⁴. Dans sa Communication sur l'intelligence artificielle pour l'Europe, la Commission n'a pas fait mention d'une telle personnalité électronique²⁶⁵. Cette omission semble indiquer qu'à ce stade, la Commission européenne n'envisage pas la création d'une personnalité juridique pour les robots ou les entités d'IA en droit de l'Union européenne²⁶⁶.

Il semblerait plus approprié de tenir les organisations de médias d'information qui décident de se servir de l'IA responsables des contenus illégaux qu'elle pourrait produire. Et si les autorités publiques souhaitaient exercer une censure sur les contenus générés automatiquement ou les bots qui contribuent au débat public, alors ces outils et les contenus qu'ils génèrent pourraient être protégés à la fois en vertu des droits à la liberté d'expression de l'organisme de média d'information et du droit de recevoir des informations des utilisateurs d'information²⁶⁷.

Au stade de la distribution de l'information, les médias d'information peuvent recourir à l'IA pour personnaliser l'offre d'information à chaque individu, ce qui peut engager le droit des personnes à recevoir de l'information. Le droit fondamental à la liberté d'expression tel que le protège l'article 10 de la CEDH comprend le droit de recevoir des informations. La Cour européenne des droits de l'homme a jugé que les médias avaient pour tâche de mettre à disposition de l'information et des idées sur des questions d'intérêt public et que le public a également le droit de les recevoir²⁶⁸. De plus, le public a le droit d'être dûment informé²⁶⁹. Le droit de recevoir des informations implique que le public devrait avoir accès à une information diverse grâce aux médias²⁷⁰. Dans le même temps, les utilisateurs d'information ne disposent pas d'un droit subjectif à recevoir de l'information de la part des

²⁶² Hage J., « Theoretical foundations for the responsibility of autonomous agents », *Artificial Intelligence and Law* 25(3), pp. 255–71, <https://doi.org/10.1007/s10506-017-9208-7>.

²⁶³ Collins R.K.L. et Skover D. M., *Robotica: Speech rights & artificial intelligence*, Cambridge University Press.

²⁶⁴ Parlement européen, Rapport contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique, paragraphe 59, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_FR.html.

²⁶⁵ Commission européenne, Communication de la Commission : L'intelligence artificielle pour l'Europe <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:52018DC0237>.

²⁶⁶ Pour une discussion détaillée sur le statut juridique de l'IA au regard du droit d'auteur, voir le chapitre 5 de cette publication.

²⁶⁷ Kaminski M.E., « Authorship, disrupted: AI authors in copyright and First Amendment law », *U.C. Davis Law Review* 51(2), pp. 589–616.

²⁶⁸ Cour européenne des droits de l'homme, *Sunday Times c. Royaume-Uni* (No. 1), paragraphe 65.

²⁶⁹ Cour européenne des droits de l'homme, *Sunday Times c. Royaume-Uni* (No. 1), paragraphe 66.

²⁷⁰ Cour européenne des droits de l'homme, *Manole et autres c. Moldova*, paragraphe 100.

médias²⁷¹. Il n'en reste pas moins que le recours à l'IA par les médias d'information peut affecter l'exercice du droit à recevoir de l'information, par exemple lorsque la personnalisation réduit la diversité des informations auxquelles les gens ont accès. Une restriction au droit de recevoir des informations résultant du comportement des médias d'information pourrait faire naître des obligations positives pour les États (voir prochaine section).

Quand les intermédiaires en ligne utilisent la personnalisation, ils peuvent également mettre en jeu les droits à la liberté d'expression de médias d'information ou de journalistes citoyens qui postent des articles d'actualité sur leurs plateformes pour atteindre un public plus large. Le gouvernement allemand fut ainsi l'un des premiers en Europe à mettre en place des protections juridiques pour la liberté des médias et la visibilité des organismes d'information au regard de la personnalisation sur les médias sociaux et moteurs de recherche. Le nouveau *Medienstaatsvertrag* allemand prévoit que les intermédiaires en ligne « ne sauraient injustement désavantager (directement ou indirectement) ou traiter différemment les fournisseurs de contenu éditorial journalistique lorsqu'ils disposent d'une influence potentiellement significative sur leur visibilité²⁷² ». Cette disposition juridique constitue une nouveauté en droit européen des médias, et bien qu'il soit trop tôt pour savoir comment la loi fonctionnera en pratique, les organisations de défense des droits numériques affirment que la nouvelle loi allemande a « une valeur symbolique importante » et que ses objectifs fondamentaux sont « louables²⁷³ ».

Enfin, les médias d'information peuvent se fonder sur l'IA pour modérer les commentaires des utilisateurs sur leurs sites web, ce qui peut mettre en jeu le droit à la liberté d'expression des individus qui postent des commentaires. Des études ont fait ressortir que les systèmes d'IA ont tendance à davantage traiter comme injurieux les messages postés sur les médias sociaux en anglais afro-américain par rapport aux messages rédigés en anglo-américain standard²⁷⁴. Si la modération des contenus par l'IA traite certains groupes sociaux de manière défavorable, cela pourrait entraîner une inégalité des opportunités de communiquer.

D'autres limites techniques de l'analyse automatisée des contenus proviennent de la difficulté qu'ont les systèmes de modération des contenus à comprendre le contexte dans

²⁷¹ Eskens S., Helberger N. et Moeller J., « Challenged by news personalisation : Five perspectives on the right to receive information », *Journal of Media Law* 9(2), pp. 259–84, <https://doi.org/10.1080/17577632.2017.1387353>.

²⁷² Helberger N., Leerssen P. et van Drunen M., « Germany proposes Europe's first diversity rules for social media platforms », *Media@LSE*, <https://blogs.lse.ac.uk/medialse/2019/05/29/germany-proposes-europes-first-diversity-rules-for-social-media-platforms/>.

²⁷³ Nelson, M., « Germany's new media treaty demands that platforms explain algorithms and stop discriminating. Can it deliver ? », *AlgorithmWatch*, <https://algorithmwatch.org/en/new-media-treaty-germany/>.

²⁷⁴ Sap M. et al., « The risk of racial bias in hate speech detection », in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 1668–78, <https://www.aclweb.org/anthology/P19-1163>. L'Assemblée parlementaire a souligné que « le recours à des ensembles de données empreints de parti pris – ou reflétant un parti pris, un préjugé ou une pratique discriminatoire du passé – constitue l'une des principales causes de la discrimination en IA », voir le rapport de l'Assemblée parlementaire: Prévenir les discriminations résultant de l'utilisation de l'intelligence artificielle (29 septembre 2020), paragraphe 43, <https://pace.coe.int/fr/files/28715/html>. L'exemple ci-dessus montre que ces risques sont également présents dans le domaine des médias.

lequel s'inscrit le commentaire d'un lecteur, du manque d'outils de traitement du langage naturel entraînés dans le domaine dans lequel ils seront déployés, et de la sous-représentation de certains groupes d'intervenants dans les données utilisées pour l'apprentissage des outils d'IA²⁷⁵. Ces limites techniques peuvent donner lieu à des faux positifs et des faux négatifs lors du passage en revue des commentaires. Comme le soulignent Llansó et ses collègues, les faux positifs peuvent entraver la liberté d'expression des individus, tandis que les faux négatifs « peuvent entraîner une défaillance du contrôle du discours de haine, du harcèlement, ou d'autres contenus indésirables qui peuvent décourager certains groupes et individus de participer en ligne²⁷⁶ ».

En principe, le droit à la liberté d'expression ne confère pas aux individus le droit de s'exprimer sur une plateforme privée. La Cour européenne des droits de l'homme a estimé que la liberté d'expression « ne donne pas la liberté de choisir un forum en vue d'exercer ce droit²⁷⁷ ». La Cour affirme néanmoins également que lorsque l'interdiction d'accéder à une propriété privée a pour effet d'empêcher tout exercice effectif de la liberté d'expression ou que la substance de ce droit s'en trouve anéantie, les États pourraient avoir l'obligation positive de protéger la jouissance du droit à la liberté d'expression en réglementant le droit de propriété²⁷⁸. De plus, comme nous l'avons évoqué dans les sections précédentes, quand les médias d'information ou les intermédiaires en ligne modèrent des commentaires, ils façonnent en pratique l'espace du débat public. La création d'un tel espace participe d'une fonction démocratique, qui s'accompagne de devoirs et responsabilités. Certains médias ont accepté ces responsabilités en ce qui concerne la modération des commentaires. Par exemple, à la suite d'études sur la manière dont la modération automatisée des contenus pouvait être discriminante ou renforcer les biais, le *New York Times* a enquêté sur la manière dont il modère automatiquement les commentaires des utilisateurs²⁷⁹.

À la suite de cette discussion sur les principes, les droits, les devoirs et les responsabilités liés à la liberté d'expression des multiples acteurs de l'environnement médiatique de l'information, il s'agit maintenant de se demander quelles sont les obligations qui pèsent sur les États en matière de liberté d'expression au regard du recours à l'IA par les médias d'information.

3.5. Obligations des États relatives à la liberté des médias

Les droits et libertés qui découlent de la CEDH y sont formulés comme des droits négatifs et les dispositions de celle-ci interdisent aux autorités publiques d'interférer avec les droits

²⁷⁵ Llansó E. *et al.*, « Artificial Intelligence, content moderation, and freedom of expression », Institut du droit de l'information, pp. 7–8, <https://www.ivir.nl/publicaties/download/AI-Llanso-Van-Hoboken-Feb-2020.pdf>.

²⁷⁶ Llansó E. *et al.*, *Ibid* p. 9.

²⁷⁷ Cour européenne des droits de l'homme, *Appleby et autres c. Royaume-Uni* [2003], 44306/98, paragraphe 47, <http://hudoc.echr.coe.int/eng?i=001-65638>.

²⁷⁸ Cour européenne des droits de l'homme, *Appleby et autres c. Royaume-Uni*, paragraphe 47.

²⁷⁹ Salganik M.J. et Lee R.C., « To apply machine learning responsibly, we use it in moderation », *NYT Open*, <https://open.nytimes.com/to-apply-machine-learning-responsibly-we-use-it-in-moderation-d001f49e0644>.

et libertés des individus. Par exemple, dans le cas de la liberté d'expression, l'article 10(1) de la Convention prévoit que tout le monde a droit à la liberté d'expression « sans qu'il puisse y avoir ingérence d'autorités publiques ». Les droits fondamentaux de la CEDH comprennent donc des obligations négatives pour les États.

Au fur et à mesure des années, la Cour européenne des droits de l'homme a admis que les droits fondamentaux issus de la CEDH pouvaient également faire naître des obligations positives pour les États. Dans les années 1960, la Cour a pour la première fois validé l'idée que les États pouvaient être tenus à des obligations positives au regard de certains droits de la Convention²⁸⁰. Il a fallu quelques années avant que la Cour affirme des obligations positives concernant le droit à la liberté d'expression. Mais dans les années 2000, la Cour a estimé que le droit à la liberté d'expression pouvait s'accompagner d'obligations positives pour les États, même dans le domaine des relations entre individus²⁸¹. Elle a notamment soutenu que les États avaient une obligation positive d'assurer que le public ait accès par le biais des médias d'information à des informations impartiales, exactes et diverses, et que les journalistes pouvaient diffuser ces informations²⁸². De plus, la Cour a observé qu'une obligation positive pesait sur les États, selon laquelle ils devaient mettre en place un cadre administratif et législatif robuste pour garantir le pluralisme du marché des médias audiovisuels²⁸³.

Dans l'affaire *Dink c. Turquie*, la Cour européenne des droits de l'homme a formulé une obligation positive particulièrement forte pour les États. Elle a estimé que les États étaient tenus de créer un environnement favorable à la participation au débat public de toutes les personnes concernées, grâce auquel celles-ci seraient à même d'exprimer sans crainte leurs idées et opinions²⁸⁴. La Cour a repris cette affirmation dans l'affaire *Khadija Ismayilova*²⁸⁵. Ces deux affaires concernaient des attaques et du harcèlement à l'encontre de journalistes. Dans l'affaire *Dink*, la Cour a jugé que l'État avait une obligation positive de protéger un journaliste contre des attaques par ceux qui se sentaient insultés par ses publications. Dans l'affaire *Khadija Ismayilova*, la Cour a estimé que l'État avait comme obligation positive d'enquêter de façon plus efficace sur les intrusions dans la vie privée d'une journaliste. Selon McGonagle, la notion d'environnement favorable pourrait avoir un impact important²⁸⁶. Mais la véritable portée de cette obligation positive pesant sur les États d'assurer un environnement favorable à la liberté d'expression reste à voir. Du point de vue

²⁸⁰ Cour européenne des droits de l'homme, *affaire relative à certains aspects du régime linguistique de l'enseignement en Belgique c. Belgique* [1968], paragraphe 27, <http://hudoc.echr.coe.int/eng?i=001-62083>.

²⁸¹ Cour européenne des droits de l'homme, *Özgür Gündem c. Turquie* [2000], 23144/93, paragraphe 43, <http://hudoc.echr.coe.int/eng?i=001-62993> ; Cour européenne des droits de l'homme, *Fuentes Bobo c. Espagne* [2000], 39293/98, paragraphe 38, <http://hudoc.echr.coe.int/eng?i=001-63608>.

²⁸² Cour européenne des droits de l'homme, *Manole et autres c. Moldavie*, paragraphe 100.

²⁸³ Cour européenne des droits de l'homme [GC], *Centro Europa 7 S.r.l. et Di Stefano c. Italie* [2012], 38433/09, paragraphe 134, <http://hudoc.echr.coe.int/eng?i=001-111396>.

²⁸⁴ Cour européenne des droits de l'homme, *Dink c. Turquie* [2010], 2668/07, 6102/08, 30079/08, 7072/09, 7124/09, paragraphe 137, <http://hudoc.echr.coe.int/eng?i=001-100383>.

²⁸⁵ Cour européenne des droits de l'homme, *Khadija Ismayilova c. Azerbaïdjan* [2019], 65286/13, 57270/14, paragraphe 158, <http://hudoc.echr.coe.int/eng?i=001-188993>.

²⁸⁶ McGonagle T., « Les obligations positives en matière de liberté d'expression : un simple potentiel ou un véritable pouvoir ? », in Andreotti O. (éd.) *Le journalisme à l'épreuve : menaces, enjeux et perspectives*, Conseil de l'Europe, pp. 9–35.

du droit positif, elle n'offre pas aujourd'hui de protection contre les risques de l'IA pour la liberté d'expression.

Bien que le récent rappel, par la Cour européenne des droits de l'homme, de l'obligation de mettre en place un environnement favorable pourrait ne pas couvrir la question du recours à l'IA par les médias d'information, ce qui précède montre que les États sont bien tenus au respect d'obligations positives susceptibles de s'appliquer dans le contexte de l'IA. Si l'utilisation de l'IA par les médias d'information réduit la diversité de l'information que les gens reçoivent, alors leurs obligations positives pourraient être déclenchées. De façon plus concrète, si la personnalisation et la modération automatisée des contenus prenaient une ampleur telle qu'elles porteraient préjudice à la diversité de l'information mise à disposition des utilisateurs, les États pourraient avoir une obligation positive de garantir qu'une information variée soit proposée aux utilisateurs sur les médias d'information recourant à l'IA. Dans ses lignes directrices sur le pluralisme des médias et la transparence de leur propriété, le Comité des Ministres du Conseil de l'Europe souligne également que les États devraient être encouragés à « veiller à ce que la plus grande diversité possible de contenus médiatiques, y compris d'intérêt général, soit accessible à tous²⁸⁷ ». Ces lignes directrices s'appliquent également à l'utilisation de l'IA par les médias d'information en ligne.

De plus, au niveau sociétal, l'utilisation de l'IA par les médias d'information et d'autres acteurs de l'environnement informationnel pourrait menacer le pluralisme des médias. Les grands médias d'information et les intermédiaires en ligne ont accès à une quantité croissante de données d'utilisateurs dont la qualité ne cesse de s'améliorer et à des technologies d'IA toujours plus puissantes, ce qui leur confère un avantage concurrentiel sur les médias d'information locaux. L'adoption de l'IA par l'industrie de l'information à plusieurs stades du processus journalistique pourrait exclure les médias locaux ou plus petits du marché, avec les risques que cela comporte de nuire au pluralisme des médias. Les États ont l'obligation positive de garantir le pluralisme des médias face à la popularité croissante de l'IA dans l'industrie de l'information, en mettant par exemple en place des règles du jeu équitables encadrant le recours par les médias d'information en ligne à l'IA et aux technologies fondées sur l'exploitation des données.

3.6. Conclusion

Ce chapitre a examiné les conséquences de l'utilisation de l'IA par les médias d'information sur le droit fondamental à la liberté d'expression tel que le protège l'article 10 de la Convention européenne des droits de l'homme. Les quatre contextes dans lesquels les médias d'information peuvent s'appuyer sur l'IA ont été passés en revue : la collecte d'informations, la production d'informations, la distribution d'informations et la modération des commentaires des utilisateurs. Le recours à l'IA par les médias d'information est en ligne avec les fonctions démocratiques des médias telles que la Cour européenne des droits

²⁸⁷ Conseil de l'Europe, Annexe à la Recommandation CM/Rec(2018)1, Lignes directrices sur le pluralisme des médias et la transparence de leur propriété.

de l'homme les reconnaît et les affirme : la distribution d'informations au public, le rôle de « chien de garde » public, et la création d'un espace de débat public.

En raison de leur rôle démocratique, les médias d'information jouissent de la liberté des médias. Celle-ci protège le recours à l'IA pour rassembler, publier, distribuer et recevoir des informations, puisque la liberté d'expression couvre tant le contenu que les moyens techniques des communications.

Les intermédiaires en ligne peuvent jouer un rôle démocratique similaire à celui des médias lorsqu'ils renforcent l'accessibilité de l'information et permettent un débat public sur leurs plateformes. La sélection, le classement et la hiérarchisation de l'information par les intermédiaires en ligne pourraient donc être couverts par la liberté d'expression ou même la liberté des médias.

Les médias d'information et les autres acteurs qui exercent ou contribuent à la liberté d'expression en s'appuyant sur l'IA sont également soumis à des devoirs et des responsabilités. Pour les médias d'information, ces devoirs et responsabilités sont inscrits dans les codes de déontologie journalistique et mis en œuvre par divers instruments d'autorégulation. Cependant, les codes de déontologie journalistique actuels n'incluent pas de lignes directrices sur l'usage de l'IA. Certains universitaires appellent donc à l'élaboration d'une déontologie du journalisme algorithmique. La notion de devoirs et responsabilités fournit également une base juridique normative permettant d'obliger les intermédiaires en ligne à développer des instruments d'autorégulation adéquats pour l'usage de l'IA lorsqu'ils contribuent à l'exercice de la liberté d'expression. Si les codes de déontologie journalistique et les autres instruments d'autorégulation continuent d'être à la traîne par rapport aux réalités de l'IA dans l'environnement médiatique de l'information, alors les États disposent d'un fondement juridique pour régler cette question, lorsque l'IA a des effets significatifs sur les droits à la liberté d'expression des utilisateurs d'information et des autres participants au débat public.

Le recours à l'IA par les médias d'information pose notamment les risques suivants à la liberté d'expression : les événements et sujets d'actualité qui ne génèrent pas de données massives peuvent être négligés par les algorithmes des systèmes automatisés de collecte d'informations et les voix des individus concernés peuvent donc passer inaperçues ; le journalisme automatisé peut produire des contenus illégaux qui enfreignent les droits et la dignité d'autres personnes ; la distribution personnalisée d'informations peut affecter le droit des utilisateurs à recevoir une information variée, ce qui forme partie intégrante de la liberté d'expression ; enfin, la modération automatisée des commentaires des utilisateurs peut défavoriser certains groupes minoritaires, ce qui pourrait entraîner une inégalité des opportunités à communiquer et prendre part au débat public.

Sur les États peuvent peser des obligations positives visant à assurer que chacun puisse effectivement exercer son droit à la liberté d'expression au regard de l'IA. Les États ont l'obligation positive de garantir que les utilisateurs d'information reçoivent une information variée et de créer un environnement favorable à la liberté d'expression. Si l'usage de l'IA par les grands organismes d'information et les intermédiaires en ligne venait à menacer la survie sur le marché des acteurs plus petits, les États pourraient être soumis à une obligation positive d'imposer des conditions équitables pour le recours à l'IA par les médias d'information.

Ces conclusions font écho aux appels des organisations de défense des droits de l'homme et des droits numériques concernant les implications de l'IA pour la liberté d'expression. Selon le Rapporteur spécial des Nations Unies sur la liberté d'opinion et d'expression, les États « devraient adopter des politiques et mesures législatives propres à créer un cyberspace favorisant le pluralisme et la diversité », ce qui inclut « de prendre des mesures destinées à promouvoir la concurrence dans le domaine de l'intelligence artificielle²⁸⁸ ». Le Représentant de l'OSCE pour la liberté des médias²⁸⁹, ainsi que Privacy International et Article 19²⁹⁰, attirent également l'attention sur les diverses menaces que fait peser l'IA sur la liberté d'expression.

Le défi qui se pose actuellement pour les médias d'information et les autorités publiques est l'élaboration de codes de déontologie journalistique, de normes d'autorégulation, et peut-être même d'une réglementation propre à maîtriser les risques que soulève l'IA pour la liberté d'expression, tout en lui permettant de contribuer au débat public, au pluralisme des médias, à la libre circulation de l'information et à d'autres objectifs sociétaux. Les principes inscrits à l'article 10 de la CEDH posent des orientations concrètes pour aider les entités publiques et privées à relever ce défi.

²⁸⁸ Kaye D., « Rapport du Rapporteur spécial sur la promotion et la protection du droit à la liberté d'expression et d'opinion », Nations Unies, paragraphe 64, <https://undocs.org/pdf?symbol=fr/A/73/348>.

²⁸⁹ Représentant de l'OSCE pour la liberté des médias, Intelligence artificielle & liberté d'expression, <https://www.osce.org/representative-on-freedom-of-media/447829?download=true>.

²⁹⁰ Privacy International et Article 19, « Privacy and freedom of expression in the age of artificial Intelligence », <http://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence>.

4. Les politiques de diversité culturelle à l'ère de l'IA

Mira Burri, Université de Lucerne

4.1. Introduction

La diversité des contenus est essentielle à un débat public dynamique, à l'inclusion culturelle et sociale ainsi qu'à la cohésion. Aussi la diversité culturelle a-t-elle été longtemps définie comme un objectif réglementaire dans les politiques nationales en matière de médias et de culture, en particulier en Europe, et le mandat visant à la protéger et à la favoriser n'a-t-il été renforcé qu'après la Convention de l'UNESCO sur la diversité culturelle de 2005²⁹¹. Si la diversité culturelle est restée un objectif clé des politiques publiques, malgré des mises en œuvre très diverses dans les différentes politiques nationales, certaines questions fondamentales sont restées sans réponse du fait de l'évolution profonde de l'environnement technologique. Deux questions critiques doivent être posées à cet égard : premièrement, dans quelle mesure les affordances du support numérique ont-elles rendu possible, mais aussi remis en cause, la diversité en ligne – tant en termes de disponibilité d'un contenu diversifié que de sa consommation réelle ? Deuxièmement, dans quelle mesure les boîtes à outils de la politique culturelle, telles qu'elles sont actuellement appliquées, sont-elles adaptées pour traiter et encourager l'engagement en faveur de contenus culturellement diversifiés ? Cette contribution montrera que les réponses à ces questions ne sont pas simples et que les décideurs politiques peuvent devoir accepter des compromis complexes, et se montrer plus innovants dans la mise en œuvre de leurs politiques culturelles – en gouvernant par le biais d'intermédiaires et de technologies.

Cette contribution s'intéressera aux affordances des médias numériques et de l'intelligence artificielle (IA) en particulier et à leurs implications pour les politiques de contenu. Elle ne s'engagera cependant pas dans les discussions plus générales sur la créativité à l'ère de l'IA²⁹², ni n'examinera l'intégration de la diversité à l'IA pour réduire les biais dans son processus de prise de décision²⁹³ ou la diversité au sein de l'industrie de l'IA²⁹⁴.

²⁹¹ Voir, par exemple, Burri M., « The UNESCO Convention on Cultural Diversity: An appraisal five years after its entry into force. » *International Journal of Cultural Property* 20, 4, pp. 357–380, novembre 2013.

²⁹² Voir dans ce contexte, Kulesz O., « Culture, machines et plateformes : l'intelligence artificielle et son impact sur la diversité des expressions culturelles », rapport pour le compte de l'UNESCO, DCE/18/12.IGC/INF.4, 2018.

²⁹³ Melendez C., « In IA, Diversity Is a Business Imperative », *The Forbes*, 14 novembre 2019.

²⁹⁴ Voir, par exemple, Paul K., « “Disastrous” » Lack of Diversity in IA Industry Perpetuates Bias », *The Guardian*, 17 avril 2019.

4.2. Comprendre l'environnement modifié de la création de contenu, de sa distribution, de son utilisation et de sa réutilisation

Les transformations survenues dans l'environnement en réseau numérique incarnées par la pénétration sociétale d'internet ont été multiples et, au fil des ans, leurs effets ont été saisis, non sans controverse, par une série d'excellentes études²⁹⁵. La place centrale des données et la prédominance de marchés de données à multiples facettes, ainsi que l'essor de l'IA, ont créé un nouveau niveau de complexité, et il est essentiel de bien comprendre la dynamique contemporaine autour du contenu pour être en mesure de concevoir des boîtes à outils culturelles adéquates. Nous nous concentrons sur les développements spécifiques qui peuvent être critiques pour la poursuite des objectifs de diversité culturelle dans ce nouvel espace et nous nous intéressons plus particulièrement aux changements survenus dans la façon dont le contenu est produit, distribué, consulté et consommé en ligne, ainsi qu'aux modifications connexes apportées aux modèles d'expérience et de participation des utilisateurs, lorsqu'elles peuvent être identifiées. Ce chapitre prend l'évolution du rôle des intermédiaires en tant que contrôleurs d'accès essentiels comme point d'entrée de cette discussion complexe.

4.2.1. Comprendre les nouveaux intermédiaires

Certains avancent, en s'appuyant parfois sur des preuves, que l'environnement numérique apporterait aux utilisateurs une abondance, une diversité et une autonomisation qui seraient impossibles dans les conditions des médias analogiques²⁹⁶. L'un des éléments essentiels qui sous-tendent ces visions positives est que les intermédiaires n'existent pas dans le cyberspace et que chacun est libre de choisir n'importe quel contenu à tout moment. Pourtant, comme le montre la pratique contemporaine des médias numériques, cette affirmation peut être erronée. En fait, il se peut qu'il y *ait* divers intermédiaires qui exercent différents types de contrôle sur les choix que nous *faisons* et sur la possibilité de choix que nous *voyons*. Nous ne parlons pas ici d'intermédiaires physiques, tels que les opérateurs de réseau ou les fournisseurs de services internet (même s'ils peuvent être très importants²⁹⁷), mais nous nous focalisons sur les contrôleurs d'accès qui existent au niveau

²⁹⁵ Voir, par exemple, Benkler Y., *The wealth of networks: How social production transforms markets and freedom*. New Haven: Yale University Press, 2006), Sunstein C. R., *Republic.com 2.0*. Princeton: Princeton University Press, 2007.

²⁹⁶ Benkler Y., Weinberger D., *Everything is miscellaneous: The power of the new digital disorder*. New York: Henry Holt, 2007, Jenkins H., *Convergence culture: Where old and new media collide*. New York: New York University Press, 2008.

²⁹⁷ Benkler Y. (2006).

des applications et du contenu – ces « facilitateurs de choix »²⁹⁸ ou « nouveaux éditeurs »²⁹⁹, qui utilisent souvent aussi les technologies de l'IA.

Pour comprendre le nouvel espace médiatique, il peut être utile de le comparer au fonctionnement des médias traditionnels. Ordinairement, dans le monde hors ligne/analogique, les rôles éditoriaux étaient concentrés sous le toit d'une seule institution. Les choix éditoriaux étaient basés sur un nombre limité de matériaux, et les produits éditoriaux étaient soumis aux limites inhérentes à chaque média, telles que les pages d'un journal imprimé ou la durée d'une émission. En règle générale, on s'adressait aussi au public visé selon un certain rythme qui influençait l'étendue et la profondeur du contenu – par exemple, par une diffusion quotidienne ou un hebdomadaire. Les décisions éditoriales prises quant au contenu et au format se répercutaient sur l'ensemble du public d'une publication ou d'un programme donné de la même manière – elles n'étaient pas adaptées à un utilisateur particulier. Selon le format, il existait également un certain équilibre entre les sujets locaux, nationaux et internationaux, qui étaient présentés de manière contextualisée et fiable. Il s'agissait des fonctions éditoriales centrales des radiodiffuseurs et autres médias traditionnels qui, dans de nombreux territoires, avaient également pour mandat spécifique de présenter des contenus locaux et nationaux ; il existait généralement des mécanismes pour superviser le respect de certaines exigences en matière de quantité et de qualité des contenus. Dans l'Union européenne, par exemple, outre l'obligation de diffuser une majorité d'œuvres européennes sur les chaînes audiovisuelles, des exigences et des obligations sont imposées au niveau national. Dans l'ensemble, ces fonctions éditoriales relativement bien définies avaient des conséquences importantes sur la production et la diffusion des connaissances³⁰⁰. Elles soutenaient également la conviction, qui est à la base de presque toutes les politiques nationales en matière de médias, que la diversité de l'offre se reflétera dans la diversité de la consommation.

Le tableau est désormais très différent car les formes de médias numériques suppriment ces limitations analogiques et entraînent « des changements fondamentaux dans la composition et la consommation des produits médiatiques »³⁰¹. Les « nouveaux éditeurs » sont multiples, désintégrés et dispersés³⁰². Les « nouveaux éditeurs » sont dirigés par l'IA et ce sont, en fin de compte, les algorithmes³⁰³ qui définissent le nouvel espace médiatique.

²⁹⁸ Helberger N., « Diversity label: Exploring the potential and limits of a transparency approach to media diversity », *Journal of Information Policy* 1. pp. 337–369, 2011, Helberger N., « Diversity by design », *Journal of Information Policy*, pp. 441–469, 2011.

²⁹⁹ Miel P. et Farris R., *News and information as digital media come of age*, Cambridge: The Berkman Center for Internet and Society, 2008, p. 27.

³⁰⁰ Weinberger D., *Too big to know*, New York: Basic Books, 2012.

³⁰¹ Miel et Farris, p. 27. Voir aussi Kleis Nielsen R., Gorwa R., et de Cock Buning M., *What can be done? Digital media policy options for strengthening European democracy*, Oxford : Rapport du Reuters Institute, 2019.

³⁰² *Ibid.*, également Latzer M., Hollnbuchner K., Just N., et Saurwein F., « The economics of algorithmic selection on the Internet » dans Bauer J. M. et Latzer M. (éd.), *Handbook on the economics of the Internet*. Cheltenham: Edward Elgar. pp. 395–425, 2016.

³⁰³ Pour une définition complète des algorithmes, voir Latzer M., et Just N., « Governance by and of algorithms on the Internet: Impact and consequences », in *Oxford Research Encyclopedia, Communication*, Oxford: Oxford University Press, 2020.

L'*agrégation* est le premier de ces éditeurs et désigne le processus d'assemblage personnalisé de différents types de contenus et leur mise à jour constante. Ce type d'éditeur personnalisé est proposé sur différentes plateformes, pour différents types de contenus – qu'il s'agisse d'informations, de divertissement ou de faits divers. Il génère automatiquement des informations personnalisées en fonction d'un profil d'utilisateur particulier et/ou d'une expérience antérieure, de façon apparemment transparente et incessante. Le mécanisme qui pilote ce flux de contenu est généralement un algorithme spécifique à la plateforme (que ce soit Facebook ou Instagram par exemple) et peut distinguer différents types de contenu³⁰⁴.

Le *social bookmarking* (ou partage de signets) est également un mécanisme qui gagne en importance lorsqu'il s'agit de mettre un contenu en avant. Ici, la communauté joue le rôle d'un éditeur grâce à différents systèmes de classement et de signets, tels que Reddit, Technorati ou Del.icio.us. Avec l'adoption à grande échelle de Twitter et d'Instagram, en particulier par les jeunes générations, l'utilisation de hashtags comme type de balise de métadonnées, permettant aux utilisateurs de créer des index dynamiques générés par l'utilisateur, a augmenté. Ces mécanismes peuvent non seulement personnaliser la consommation de médias, mais aussi réussir à attirer l'attention de grands groupes³⁰⁵. Cela peut être vrai pour les campagnes politiques, mais aussi pour la mobilisation de l'attention des consommateurs dans les campagnes de marketing.

Enfin, en tant qu'intermédiaire numérique, la *recherche* est aujourd'hui absolument essentielle. Elle est souvent le point de départ de la majorité des expériences en ligne et constitue le principal moteur de trafic de la plupart des sites web. S'il n'est pas indexé et ne peut être trouvé sur internet, un contenu devient tout simplement inexistant³⁰⁶. Là encore, les recherches sont généralement dirigées par des algorithmes propriétaires et l'activité est fortement concentrée autour de quelques fournisseurs, Google se démarquant clairement de ses concurrents.

³⁰⁴ Les algorithmes combinent souvent différents mécanismes et sont commandés par différents facteurs : (1) la popularité générale de l'article auprès de l'ensemble des utilisateurs est l'approche la plus simple, par laquelle les utilisateurs obtiennent la même recommandation, ce qui a pour résultat final que les articles populaires deviennent encore plus populaires et que les articles impopulaires disparaissent ; (2) le filtrage sémantique recommande les articles qui correspondent à l'article actuellement utilisé ou aux articles précédemment utilisés par le même utilisateur selon un certain nombre de critères prédéfinis (tels que les sujets, l'auteur ou la source d'un article) ; (3) le filtrage collaboratif ou le filtrage des informations sociales est un générateur de recommandations automatisé de « bouche à oreille » – les articles sont recommandés à un utilisateur en fonction des valeurs attribuées par d'autres personnes ayant des goûts similaires. Ces méthodes sont généralement appliquées sous des formes hybrides, incluant également d'autres méthodes comme la pondération des articles en fonction de leur récence ou le push de contenu ayant des caractéristiques spécifiques comme le contenu payant. Les plateformes ont également accumulé au fil des ans de grandes quantités de précieuses données basées sur le comportement passé et peuvent en outre appliquer des données utilisateur telles que l'âge ou la localisation pour calibrer le flux de contenu. Voir Bozdag E., « Bias in algorithmic filtering and personalization », in *Ethics and Information Technology* 15. pp. 209–227, 2013.

³⁰⁵ Miel et Farris, p. 30.

³⁰⁶ Conseil de l'Europe, Projet de recommandation sur la protection des droits de l'homme dans le contexte des moteurs de recherche, Strasbourg, 11 mars 2010.

4.2.2. Implications d'agents éditoriaux dirigés par l'IA

Grâce à tous ces différents mécanismes, le réseau fonctionne comme un éditeur multicanal et un intermédiaire important dans la chaîne de valeur du contenu – il remplace les médias traditionnels dans leur rôle d'« intermédiaire d'intérêt général »³⁰⁷. Parmi les aspects positifs, il a été suggéré que « l'environnement médiatique en réseau en tant qu'esprit social virtuel [...] produit quelque chose de plus riche, de plus représentatif et de plus ouvert aux idées que le modèle de médias de masse de haut en bas du passé »³⁰⁸. Bien que nous ne devions pas sous-estimer les affordances des plateformes numériques et les processus de communication, de participation et d'engagement qu'elles permettent, du moins jusqu'à présent, il existe une profonde incertitude, voire un doute croissant, quant à la capacité de ce mécanisme auto-organisé à identifier de manière fiable les informations importantes³⁰⁹. Il existe également une dose de scepticisme quant à son impact sur la diversité et la qualité des contenus, et sur la capacité des utilisateurs à trouver et à accéder à des contenus diversifiés et fiables.

En réfléchissant aux fonctions sociétales des médias et à l'objectif de diversité culturelle dans le contexte de la discussion de ce chapitre, il se pourrait que cet environnement complexe présente certains dangers de réduction de la diversité et de fragmentation du débat public³¹⁰. Tout d'abord, nous devons reconnaître les interférences possibles avec l'autonomie et la liberté de choix individuelles des utilisateurs. Comme le soutiennent Latzer *et al.*, si le filtrage réduit les coûts de recherche et d'information et facilite l'orientation sociale³¹¹, il peut être « compromis par la production de risques sociaux, notamment des menaces aux droits et libertés fondamentaux ainsi que par des répercussions sur la médiation des réalités et le développement futur des personnes »³¹². La deuxième préoccupation dans ce contexte concerne l'impact de la production et de la consommation de médias personnalisés. Au sens premier, on observe une tendance récente à la production de contenu algorithmique, les algorithmes orientant la prise de décision dans les organisations médiatiques en prédisant les modèles de consommation et les préférences du public³¹³. Si dans certains domaines, cela peut être considéré comme bénéfique pour donner au public ce qu'il veut, cela peut s'avérer, dans d'autres domaines comme pour l'information, très problématique car les informations et les actualités locales sont ajustées aux variables démographiques, sociales et politiques de communautés spécifiques³¹⁴. Il convient également de mentionner les « fermes de contenu » qui, sur la base des données des moteurs de recherche (telles que les termes de recherche populaires, les ventes de mots publicitaires et le contenu réellement disponible), produisent du

³⁰⁷ Sunstein (2007).

³⁰⁸ Miel et Farris, p. 30.

³⁰⁹ *Ibid.*

³¹⁰ Voir, par exemple, Sunstein C. R. *Going to extremes: How like minds unite and divide*, Oxford: Oxford University Press, 2009., Pariser E., *The filter bubble: What the Internet is hiding from you*, Londres : Viking, 2011.

³¹¹ Latzer *et al.*

³¹² *Ibid.*, pp. 29–30.

³¹³ Napoli P. M., « On automation in media industries: Integrating algorithmic media production into media industries scholarship » in *Media Industries Journal* 1. pp. 33–38, 2014 ; également Saurwein F., Just N., et Latzer M., « Governance of algorithms: Options and limitations », *info* 17. pp. 35–49, 2015.

³¹⁴ Napoli, *ibid.*, p. 34.

contenu rapidement et à moindre coût afin de répondre à cette demande. Une telle création de contenu s'assimile à de la marchandisation et peut nuire à toute fonction d'intérêt public des médias, y compris dans la sphère culturelle.

Au sens de la consommation de médias, la personnalisation du régime médiatique, basée sur un profil spécifique ou une expérience antérieure, « favorise un contenu géographiquement proche ainsi que socialement et conceptuellement familier »³¹⁵ ... « Cela maintient les utilisateurs dans des limites familières, alimentant leur curiosité en leur donnant toujours la même chose. Lorsqu'ils recherchent de nouveaux contenus ou de nouvelles informations, les opinions existantes sont renforcées, éliminant ainsi progressivement les points de vue contradictoires »³¹⁶. On peut bien sûr faire valoir que c'était également le cas avec les médias traditionnels : les gens sont naturellement attirés par le contenu qu'ils ont aimé dans le passé – la principale différence dans l'espace actuel est que les utilisateurs *ne voient que ce* contenu, et leur choix actif est ainsi diminué ou manipulé. Hoffman *et al.* soutiennent que les médias sociaux ne font qu'exacerber cet effet en combinant deux dimensions de l'homophilie : la similitude des pairs et du contenu³¹⁷. Dans ce contexte, nous devons garder à l'esprit que, malgré une légère diminution de leur utilisation comme point d'accès au contenu et l'existence de variations d'un pays à l'autre³¹⁸, les sites de réseaux sociaux restent d'importants contrôleurs d'accès. Cela renforce l'effet d'homophilie, tout en illustrant clairement le pouvoir de quelques acteurs et l'impact profond de leurs décisions – par exemple, lorsque Facebook a changé son algorithme en 2018 et déclassé les actualités, la consommation d'actualités a automatiquement diminué³¹⁹.

Il convient également de souligner la commercialisation des plateformes et l'augmentation sensible des « fausses nouvelles » à caractère commercial ou politique³²⁰. Malgré le léger glissement vers des modèles de paiement par les lecteurs pour les actualités, il est utile de rappeler que la grande majorité de la consommation en ligne se fait encore par le biais de sites web gratuits, largement soutenus par la publicité. Bien qu'une partie du contenu agrégé soit tirée des médias traditionnels³²¹, ce qui permet de balayer certaines des critiques habituelles selon lesquelles les agrégateurs amplifient l'impact des sources non traditionnelles peu fiables, il n'en reste pas moins vrai que le contenu n'est pas rendu plus abondant, il est simplement davantage dispersé – en ce sens, nous n'avons pas un contenu plus abondant et plus diversifié, mais simplement plus de la

³¹⁵ Hoffman C. P., Lutz C., Meckel M., et Ranzini G., « Diversity by choice: Applying a social cognitive perspective to the role of public service media in the digital age », *International Journal of Communication*, 9, 2015, pp. 1360–1381.

³¹⁶ *Ibid.*

³¹⁷ *Ibid.*

³¹⁸ Pour des analyses par pays, voir Reuters Institute for the Study of Journalism, *Digital News Report 2018*, Oxford, 2018.

³¹⁹ Reuters Institute for the Study of Journalism (2018) ; également Tucker J. A. *et al.* *Social media, political polarization, and political disinformation: A review of the scientific literature*, préparé pour la Hewlett Foundation, mars 2018.

³²⁰ Reuters Institute for the Study of Journalism (2018) ; Commission européenne, Lutter contre la désinformation en ligne : une approche européenne, COM(2018) 236 final, 26 avril 2018.

³²¹ Reuters Institute for the Study of Journalism (2018). Les agrégateurs peuvent être quelque peu limités par le droit d'auteur, voir *Associated Press v. Meltwater U.S. Holdings, Inc.* 931 F. Supp. 2d 537, 537 (S.D.N.Y. 2013) et des initiatives plus récentes dans le domaine de la législation européenne sur le droit d'auteur.

même chose. Néanmoins, il est juste de noter que les médias traditionnels ont réagi à l'agrégation rendue possible par la technologie et offrent beaucoup plus de contenu en ligne que dans leurs versions imprimées ou diffusées. En ce qui concerne plus particulièrement l'information, le Reuters Institute for the Study of Journalism a constaté que les organismes de presse privés investissaient massivement dans les médias sociaux et faisaient état d'un trafic important, d'une portée hors site et/ou de nouveaux abonnés numériques³²². Si cela peut permettre d'accéder à une variété de contenus sur un plus grand nombre de plateformes, ce qui est également intéressant pour les jeunes, deux inconvénients doivent être soulignés : le premier concerne la dépendance presque totale des organisations médiatiques à l'égard de Facebook, d'où un certain « risque de plateforme » ; le second est que la façon dont les organes d'information traditionnels du secteur privé abordent les médias sociaux est fortement influencée par des modèles économiques dépendant du parcours emprunté, orientés vers la publicité et les abonnements, ou un mélange des deux³²³. Une fois encore, cela pourrait ne pas conduire à une offre durable de contenus locaux, régionaux et nationaux diversifiés³²⁴. Dans l'ensemble, bien que la quantité de contenu augmente, le contenu local, régional et national peut diminuer et devenir très difficile à trouver parce qu'il est – ou devient – marginalisé sur les plateformes en ligne.

En ce qui concerne le rôle d'intermédiaire joué par les moteurs de recherche, il peut être généralement dans l'intérêt à long terme des fournisseurs de services de recherche de répondre aux besoins de leurs utilisateurs – à la fois en tant que consommateurs et en tant que citoyens. Au vu des recherches menées par l'Ofcom (l'autorité britannique de régulation des communications et des médias) qui suggèrent que la demande de contenu de service public national reste forte et il devrait donc continuer à être dans l'intérêt des fournisseurs de services de recherche de s'assurer que leurs résultats accordent la place prééminente qu'il mérite à ce contenu³²⁵. Une étude comparative récente a également révélé que les personnes qui trouvent des informations via les moteurs de recherche utilisent en moyenne davantage de sources d'informations en ligne, sont plus susceptibles d'utiliser des sources d'informations en ligne de gauche comme de droite, et ont des répertoires d'informations plus équilibrés³²⁶. Cela dit, et comme mentionné précédemment, les résultats de recherche

³²² Il identifie trois objectifs stratégiques principaux qui déterminent les différentes manières dont les organes d'information abordent les médias sociaux : (1) en stimulant le trafic sur le site par des recommandations ; (2) en stimulant la portée hors site par des formats natifs et des contenus dispersés ; (3) en stimulant les ventes d'abonnements numériques, souvent en partie par des contenus publicitaires sur Facebook.

³²³ Cornia A., Sehl A., Levy D. A., et Nielsen R. K., *Private sector news, social media distribution, and algorithm change*, Oxford: Reuters Institute for the Study of Journalism, 2018.

³²⁴ Une étude des médias locaux américains a montré, par exemple, que seulement 17 % des informations fournies à une communauté étaient véritablement locales – c'est-à-dire qu'elles concernaient ou avaient eu lieu dans la municipalité ; moins de la moitié (43 %) des informations étaient originales – c'est-à-dire qu'elles étaient effectivement produites par le média local. Voir Napoli P. M., Weber M., McCollough K., et Wang Q., *Assessing local journalism: News deserts, journalism divides, and the determinants of the robustness of local news*. News Measures Research Project, août 2018.

³²⁵ Ofcom, *Ofcom's Second Public Service Broadcasting Review, Phase Two: Preparing for the Digital Future*, London: Ofcom, 2008, paragraphe 5.60.

³²⁶ Les auteurs font référence à un phénomène de « sérendipité automatisée », qui conduit les gens vers des sources qu'ils n'auraient pas utilisées autrement. Voir Fletcher R. et Nielsen R. K., « Automated serendipity » in *Digital Journalism* 6. pp. 976–989, 2018.

sont générés de manière algorithmique et attribuent automatiquement une pertinence à certaines catégories d'information. La sélection automatisée est également sujette à des manipulations à l'aide d'une série de techniques d'optimisation des moteurs de recherche, grâce auxquelles les contenus sponsorisés ou autres gagnent en visibilité et attirent davantage l'attention³²⁷.

En conclusion de cette section, qui n'offre qu'un aperçu de la complexité de notre environnement médiatique contemporain, il convient de souligner sa fluidité et l'incertitude qui en découle quant à son impact, pour ce qui est de l'abondance et de la diversité de contenu et des conditions de la liberté d'expression. D'une part, il existe un discours dans la littérature qui, sous différentes étiquettes telles que « bulles de filtres »³²⁸ ou « chambres d'écho »³²⁹, met en évidence les risques du régime médiatique sur mesure actuel qui pourrait conduire à une fragmentation du débat public et à une possible polarisation des points de vue³³⁰. D'autre part, nous ne sommes pas sûrs de savoir dans quelle mesure cela est vrai. Un rapport transnational de 2017 constatait, par exemple, que, même si la recherche joue un rôle majeur dans la formation de l'opinion, elle doit être considérée dans un contexte de médias multiples et n'est pas déterministe³³¹. L'étude de la « sérendipité automatisée », qui dénote un phénomène par lequel les utilisateurs sont attirés vers des sources qu'ils n'auraient pas consultées autrement, réduit également les craintes suscitées par les « chambres d'écho »³³². Dans le même contexte, il convient de noter que les outils actuellement utilisés pour suivre la fragmentation nous en disent étonnamment peu sur la fidélité du public et sur la manière dont l'attention du public passe d'un média à l'autre³³³. Webster et Ksiazek trouvent par exemple peu de preuves indiquant que le public est composé de fidèles dévoués³³⁴. « De plus, les mesures de l'exposition, aussi précises soient-elles, ne peuvent pas nous dire comment le contenu affecte les gens. Il se peut que même de modestes périodes d'exposition à des discours de haine ou à d'autres médias obscurs aient des effets puissants sur ceux qui les recherchent »³³⁵, ou il se peut que l'effet global soit équilibré par d'autres composantes du régime médiatique³³⁶. En ce sens, nous ne devrions pas nous concentrer sur des instantanés mais examiner la dynamique et suivre l'évolution dans le temps³³⁷.

³²⁷ Voir, par exemple, Bradshaw S., « Disinformation optimised: Gaming search engine algorithms to amplify junk news », in *Internet Policy Review* 8. pp. 1–24, 2019.

³²⁸ Pariser (2011).

³²⁹ Sunstein C. R., *Infotopia: How many minds produce knowledge*. Oxford : Oxford University Press, 2006.

³³⁰ Voir aussi le Groupe de haut niveau sur la liberté et le pluralisme des médias, *Des médias libres et pluralistes pour soutenir la démocratie européenne*. Rapport préparé pour la Commission européenne p. 27, janvier 2013.

³³¹ Dutton W. H. et al « Search and politics: The uses and impacts of search in Britain, France, Germany, Italy, Poland, Spain, and the United States », document de travail du Quello Center No 5 pp. 1-17, 18 mai 2017.

³³² Fletcher et Kleis Nielsen (2018).

³³³ Webster J.G. et Ksiazek T. B., « The dynamics of audience fragmentation: Public attention in an age of digital media », *Journal of Communication* 62, pp. 39–56, 2012.

³³⁴ *Ibid.*, p. 40.

³³⁵ *Ibid.*, p. 51.

³³⁶ Pour des constatations très intéressantes, voir Pew Research Center, « Political polarization and media habits: From Fox News to Facebook, how liberals and conservatives keep up with politics », Washington, DC : Pew Research, 2014.

³³⁷ Pour connaître les nouvelles tendances en matière de consommation de médias, voir le Reuters Institute for the Study of Journalism (2018).

4.3. Pistes d'action possibles : de nouveaux outils pour aborder et impliquer les intermédiaires numériques

En brossant le tableau du paysage médiatique transformé et évolutif présenté ci-dessus, nous observons la complexité des nouveaux processus « éditoriaux » et la difficulté pour les individus à naviguer dans cet espace de contenu potentiellement riche mais dispersé. Nous identifions également certains risques potentiels liés à une consommation de contenu sur mesure et à une polarisation des points de vue dans cet environnement, car les nouveaux intermédiaires dirigent l'offre et la demande de manière algorithmique en sélectionnant les éléments d'information et en leur attribuant une pertinence³³⁸. Dans ce contexte, on peut envisager deux canaux viables pour introduire des mesures en faveur de la diversité culturelle : le premier consiste à s'attaquer à l'environnement émergent et à gérer les algorithmes comme des contrôleurs d'accès critiques, ces derniers étant restés jusqu'à présent largement non réglementés, en particulier à des fins de politique culturelle ; le second consiste à utiliser les nouveaux intermédiaires comme un outil pour promouvoir l'exposition à la diversité culturelle, dans le sens d'une « gouvernance par le biais des intermédiaires ».

4.3.1. La gouvernance des algorithmes

La question de la gouvernance des algorithmes donne lieu à un débat plus générique, pas nécessairement lié à la politique culturelle, qui a trait à l'observation du fait que les intermédiaires, en particulier ceux qui sont dirigés par les algorithmes, ont acquis un rôle critique dans l'espace en ligne, lequel est, en ce sens, désormais régi *par* les algorithmes³³⁹. Cette discussion est étroitement liée à celle portant sur les moyens appropriés d'aborder ce nouveau pouvoir – à savoir la gouvernance *des* algorithmes³⁴⁰. Les questions de protection de la vie privée sont particulièrement importantes dans ce dernier contexte³⁴¹, mais aussi les questions relatives à l'application du droit d'auteur par le biais des intermédiaires, comme l'illustrent la dernière réforme du droit d'auteur de l'UE et la discussion autour de l'article 17 de la Directive sur le droit d'auteur dans le marché unique numérique³⁴². Latzer *et al.* identifient neuf catégories de risques découlant de la sélection algorithmique dont il serait bon de tenir compte : (1) la manipulation ; (2) la diminution de la variété, les

³³⁸ Saurwein *et al.* (2015), p. 35.

³³⁹ Pour une excellente analyse et revue de la littérature, voir Saurwein *et al.* (2015).

³⁴⁰ Saurwein *et al.* (2015).

³⁴¹ Par exemple en ce qui concerne le droit à l'oubli, qui est désormais inscrit dans le régime de protection des données de l'UE du Règlement général sur la protection des données.

³⁴² Voir la Directive 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, JO L (2019) 130/92, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0790>. Voir aussi Montagnani M. L. et Trapova A. Y., « Safe harbours in deep waters: A new emerging liability regime for Internet intermediaries in the digital single market », *International Journal of Law and Information Technology* 26, pp. 294–310, 2018 ; Senftleben M., « Institutionalized algorithmic enforcement – The pros and cons of the EU approach to UGC platform liability », *Florida International University Law Review* 14, 2020.

chambres d'écho et les biais ; (3) les contraintes à la liberté d'expression ; (4) la surveillance et les menaces pour la protection des données et le respect de la vie privée ; (5) la discrimination sociale ; (6) la violation des droits de propriété intellectuelle ; (7) l'abus de pouvoir de marché ; (8) les effets sur les capacités cognitives ; et (9) l'hétéronomie croissante, la perte de la souveraineté humaine et la contrôlabilité de la technologie³⁴³. Nous nous sommes particulièrement intéressés aux points (2), (3) et (9) ci-dessus, car ils sont directement liés aux objectifs fondamentaux de la diversité culturelle poursuivis dans le domaine des médias.

Saurwein *et al.* présentent une analyse minutieuse des différentes options de gouvernance susceptibles de juguler ces risques : des interventions classiques d'injonction et de contrôle à la régulation par le marché, en passant par diverses solutions d'autorégulation et de corégulation³⁴⁴. Toutefois, les auteurs notent également qu'il n'y a guère eu jusqu'à présent d'outils conçus pour traiter les risques de biais, d'hétéronomie et les effets sur les capacités cognitives. Il est vrai, dans le cadre spécifique de notre discussion, que la plupart de ces plateformes intermédiaires ne relèveront pas du champ d'application réglementaire des régimes de médias actuels. Napoli a fait valoir dans ce contexte que nous devrions commencer à considérer les algorithmes comme une forme particulière d'institution médiatique³⁴⁵. Il estime que les algorithmes devraient être soumis à des types de réglementation restrictifs – à savoir l'interdiction de certains types d'activités par les opérateurs de plateformes ou de contenus sur ces plateformes, afin de protéger la vie privée et de contrer la violence graphique et les discours de haine. Napoli suggère que, compte tenu du rôle crucial que jouent ces nouveaux intermédiaires, nous devrions développer des « approches positives dans l'intérêt public »³⁴⁶, comme nous l'avons fait pour les médias électroniques traditionnels, fondées sur des principes établis en matière de politique médiatique, tels que la pluralité, la diversité et le localisme – en prescrivant par exemple certaines quantités ou certains types de contenu³⁴⁷. Cela pourrait, d'une part, remédier au manque de contenus diversifiés, fiables et locaux qui, malgré l'augmentation du nombre de contenus en ligne, semblent rares et, d'autre part, accroître la possibilité pour les utilisateurs d'accéder à ces contenus³⁴⁸.

Il convient de noter dans ce contexte que les discussions sur la réglementation des algorithmes, bien que relativement récentes, évoluent rapidement. À l'ère des données massives, des dangers liés à la protection des données à caractère personnel et de la prolifération des fausses nouvelles, le sujet a attiré l'attention des hommes politiques et du

³⁴³ Saurwein *et al.* (2015), p. 37.

³⁴⁴ Saurwein *et al.* (2015).

³⁴⁵ Napoli P. M., « The algorithm as institution: Toward a theoretical framework for automated media production and consumption », article présenté à la conférence Media in Transition, Massachusetts Institute of Technology, mai 2013 ; Napoli P. M., « Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. Document de travail Media + the Public Interest Initiative, 2014.

³⁴⁶ Napoli (2014), *ibid.*

³⁴⁷ Napoli P. M. « Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers », *Telecommunications Policy* 39. pp. 751–760, 2015.

³⁴⁸ Voir Napoli P. M. « Re-evaluating the long tail: Implications for audiovisual diversity on the Internet », in Albornoz L. A. and Garcia Leiva M. T. (éd.), *Audiovisual industries and diversity: Economics and policies in the digital era*. Abingdon : Routledge, 2019, au chapitre 5 ; Napoli P. M., *Social media and the public interest: Media regulation in the disinformation age*, New York : Columbia University Press, 2019.

grand public, et la nécessité d'agir a été reconnue³⁴⁹. Toutefois, la forme de l'action reste à définir. Jusqu'à présent, les mesures prises ne semblent concerner que les domaines de l'autorégulation et des approches réglementaires souples, car une intervention en force pourrait non seulement entraver l'innovation des plateformes et des utilisateurs, mais aussi faire échouer l'objectif même de promotion de la liberté d'expression sous ses formes active et passive. Dans ce contexte, et pour tenter de concevoir des outils de gouvernance appropriés et tournés vers l'avenir, il convient d'étudier attentivement l'expérience acquise jusqu'à présent dans le domaine de la lutte contre la désinformation en ligne. D'une part, nous devons examiner dans quelle mesure les entreprises ont répondu à la sensibilisation accrue du public et aux exigences de confiance et de qualité des utilisateurs, et dans quelle mesure différentes solutions techniques (par exemple en ajustant les algorithmes) et autres solutions (par exemple en travaillant avec les utilisateurs et d'autres organisations) ont effectivement contribué à l'objectif – dans ce cas : limiter la quantité et la diffusion des fausses nouvelles³⁵⁰. Les efforts de Facebook, à la suite de la campagne présidentielle américaine de 2016, peuvent en être un bon exemple. En réaction à diverses accusations, Facebook a soutenu un certain nombre d'initiatives – par exemple, il a diffusé des outils éducatifs pour la maîtrise de l'information, a lancé le Facebook Journalism Project et a rejoint la News Integrity Initiative avec plusieurs partenaires universitaires et médiatiques qui s'efforcent de favoriser des communautés engagées et des médias plus inclusifs, tout en cherchant à mieux comprendre la désinformation. Concrètement et afin de réduire la diffusion de « fausses nouvelles », Facebook a conclu des partenariats avec une quarantaine d'organisations – médiatiques tierces telles que Snopes, PolitiFact, Associated Press et FactCheck.org. Celles-ci s'efforcent de vérifier les faits des reportages partagés et de les identifier par un label « *disputed* » (contesté) s'ils ne passent avec succès un contrôle des faits. Facebook a également installé un bouton « plus d'infos » qui permet aux utilisateurs d'obtenir des informations supplémentaires sur les articles de leurs flux d'informations³⁵¹. Malgré ces efforts importants, les recherches et les preuves anecdotiques suggèrent que les pratiques de Facebook sont peut-être encore insuffisantes pour garantir un espace médiatique « sain »³⁵².

Le rôle des gouvernements, de la société civile et des autres organisations d'utilisateurs devrait également être pris en compte. Un bon exemple récent d'efforts coordonnés et de collaboration sur plusieurs fronts est l'approche européenne en matière de fausses nouvelles³⁵³. La Commission européenne souscrit à l'amélioration de la

³⁴⁹ Voir, par exemple, Balkin J., « Free speech in the algorithmic society: Big data, private governance, and new school speech regulation », *UC Davies Law Review* 51. pp. 1149–1210, 2018. De nombreuses actions ont également été menées au sein de l'UE en matière de protection des données et de fausses nouvelles, par exemple grâce à l'adoption du RGPD. Voir aussi Commission européenne (2018).

³⁵⁰ Reuters Institute for the Study of Journalism (2018).

³⁵¹ Le bouton « informations complémentaires » pour les articles d'actualité qui apparaissent dans le fil d'actualité permet aux utilisateurs de cliquer pour consulter : (1) des informations tirées de la page Wikipédia concernant l'éditeur ; (2) d'autres articles récemment publiés par l'éditeur ; (3) une carte thermique indiquant où dans le monde l'article est partagé et quels amis Facebook de l'utilisateur l'ont partagé.

³⁵² Levin S., « They don't care: Facebook fact-checking in disarray as journalists push to cut ties », *The Guardian*, 13 décembre 2018. Pour une analyse plus approfondie, voir Saurwein F. et Spencer-Smith C., « Combating disinformation on social media: Multilevel governance and distributed accountability in Europe », *Digital Journalism*, 2020.

³⁵³ Commission européenne (2018).



transparence des informations diffusées, de leur diversité et de leur crédibilité, ainsi qu'à un effort visant à élaborer des solutions inclusives à long terme à cet effet. Elle a, entre autres, convoqué un forum multipartite afin de fournir un cadre pour une coopération efficace entre les parties prenantes concernées, y compris les plateformes en ligne, le secteur de la publicité et les principaux annonceurs, et d'obtenir un engagement en vue de coordonner et d'intensifier les efforts pour lutter contre la désinformation. Le premier résultat du forum a été un code de bonnes pratiques contre la désinformation à l'échelle de l'UE. Adopté en septembre 2018, le code définit des normes d'autorégulation pour lutter contre la désinformation ; il vise à atteindre les objectifs de la Commission en fixant un large éventail d'engagements, allant de la transparence de la publicité à caractère politique à la fermeture des faux comptes et à la réduction des possibilités de monétisation pour les fournisseurs de désinformation. Le code comprend également une annexe identifiant les meilleures pratiques que les signataires s'engagent à appliquer pour mettre en œuvre les engagements du code³⁵⁴. Des mesures plus décisives en matière de responsabilisation, voire une évolution vers des approches de corégulation, pourraient toutefois être nécessaires³⁵⁵, comme le montre le problème aigu des « fausses nouvelles » concernant la pandémie de Covid-19³⁵⁶.

4.3.2. La gouvernance par le biais des algorithmes

Dans la perspective d'une boîte à outils de la diversité culturelle plus ciblée, on peut envisager d'approuver de nouvelles formes d'intelligence éditoriale³⁵⁷, comme une sorte de médiation d'intérêt public de l'espace numérique qui cherche à accroître la visibilité ainsi que la facilité de découverte et d'utilisation de types de contenus discrets³⁵⁸. Nous pouvons également envisager des outils qui encouragent la diversité de l'exposition, c'est-à-dire la consommation réelle de contenus diversifiés³⁵⁹.

³⁵⁴ Pour tous les documents, voir <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>. Le dernier rapport du Groupe des régulateurs européens des services de médias audiovisuels (ERGA) sur la désinformation montre certaines limites dans les engagements pris par les plateformes dans le cadre du code ; voir ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice, 2020, <https://erga-online.eu/?p=732>.

³⁵⁵ Saurwein et Spencer-Smith (2020).

³⁵⁶ Voir, par exemple, « Social media firms fail to act on Covid-19 fake news », *BBC News*, 4 juin 2020.

³⁵⁷ Nous n'aborderons pas ici les politiques d'éducation aux médias, qui peuvent également être importantes du point de vue de l'utilisateur. Voir Helberger (2011), p. 357 ; Burri M., « The global digital divide as impeded access to content », dans Burri M., et Cottier T. (éd.), *Trade governance in the digital age*. Cambridge : Cambridge University Press. pp. 396-420, 2012 ; Groupe de haut niveau sur la liberté et le pluralisme des médias, « Des médias libres et pluralistes pour soutenir la démocratie européenne », rapport préparé pour la Commission européenne, janvier 2013.

³⁵⁸ Miel et Farris, p. 3 ; également Goodman E. P., « Public media 2.0 », dans Schejter A. M. (éd.), *And communications for all: A public policy agenda for a new administration*. Lanham, MD : Lexington Books, pp. 263-280, 2009 ; Webster J. G., « User information regimes: How social media shape patterns of consumption », *Northwestern University Law Review* 104, pp. 593–612, 2010.

³⁵⁹ Voir Helberger N., « Media diversity from the user's perspective: An introduction », *Journal of Information Policy* 1, pp. 241–245, 2011 ; Napoli P. M., « Exposure diversity reconsidered », *Journal of Information Policy* 1, pp. 246–259, 2011.



Ce projet n'est pas complètement nouveau ni insolite. Le cadre européen pour les médias, en vertu de la Directive Services de médias audiovisuels (Directive SMA)³⁶⁰, suggère que la promotion des œuvres européennes peut être liée à une meilleure « mise en valeur » de ces œuvres³⁶¹. D'après la consultation des autorités de régulation en 2013³⁶², il apparaît que beaucoup d'entre elles étaient favorables aux outils de mise en valeur, tout en étant sceptiques quant à la promotion par le partage des actifs dans les catalogues. Beaucoup considèrent cette mesure comme la plus efficace (également parce qu'elle concerne une consommation réelle plus élevée d'œuvres européennes) et la moins lourde pour les opérateurs³⁶³. En Europe, les radiodiffuseurs de service public ont eu le privilège d'occuper les premiers créneaux horaires des guides électroniques des programmes (GEP) et ont ainsi bénéficié d'une « mise en valeur particulière »³⁶⁴. Foster et Broughton montrent que les GEP ont été un outil important permettant aux consommateurs de trouver et de sélectionner des programmes, et il est prouvé que les chaînes dont les créneaux sont situés près du haut de chaque section d'un GEP ont été davantage sélectionnées par les utilisateurs et ainsi avantagées par rapport à celles situées plus bas³⁶⁵. « Cette approche [consistant à « pousser » les gens à faire les choix que nous espérons qu'ils feront dans leur propre intérêt et dans celui de la société en général] a jusqu'à présent plutôt bien fonctionné »³⁶⁶. Des preuves récentes confirment que son positionnement dans les GEP est susceptible d'avoir un impact significatif sur les résultats d'une chaîne³⁶⁷. Ces « coups de pouce », bien que destinés aux médias commerciaux, ont également fonctionné avec les télécommandes des téléviseurs SMART, les grands acteurs en ligne, tels que Netflix, YouTube et Google Play, apparaissant comme des boutons permettant un accès direct.

³⁶⁰ La Directive SMA a été révisée pour la dernière fois en 2018 : voir Directive (UE) 2018/1808 du Parlement européen et du Conseil du 14 novembre 2018 visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à la fourniture de services de médias audiovisuels (Directive Services de médias audiovisuels), compte tenu de l'évolution des réalités du marché, JO L (2018) 303/69.

³⁶¹ Voir les dernières recommandations de l'UE : Communication de la Commission - Lignes directrices en vertu de l'article 13, paragraphe 7, de la Directive Services de médias audiovisuels concernant le calcul de la part des œuvres européennes dans les catalogues des services de médias audiovisuels à la demande et la définition d'une faible audience et d'un chiffre d'affaires peu élevé 2020/C 223/03, JO C (2020) 223/10.

³⁶² Voir, par exemple, Observatoire européen de l'audiovisuel, IRIS Spécial : La vidéo à la demande et la promotion des œuvres européennes. Strasbourg : Observatoire européen de l'audiovisuel, 2013.

³⁶³ Commission européenne (2014), p. 6.

³⁶⁴ Les GEP ont été réglementés au niveau de l'UE par la directive Accès (Directive 2002/19/CE du Parlement européen et du Conseil du 7 mars 2002 relative à l'accès aux réseaux de communications électroniques et aux ressources associées, ainsi qu'à leur interconnexion, JO L 108/7 du 24 avril 2002, modifiée par la Directive 2009/140/CE du Parlement européen et du Conseil du 25 novembre 2009, JO L 337/37 du 18 décembre 2009). La transposition de la directive diffère – par exemple, la réglementation britannique permet un traitement préférentiel des chaînes de service public, alors qu'en Allemagne, la réglementation des GEP est basée sur l'égalité de traitement des chaînes publiques et commerciales dans les listes des GEP.

³⁶⁵ Foster R. et Broughton T., « PSB prominence in a converged media world », rapport commandé par la BBC, Londres : Communications Chambers, p. 12, 2012. Parmi les autres facteurs qui influencent la sélection, citons le fait d'avoir un numéro de chaîne GEP mémorisable et de se trouver à côté d'une autre chaîne populaire.

³⁶⁶ Foster et Broughton (2012), p. 13-14. Cela a été confirmé par un rapport plus récent.

³⁶⁷ Ofcom, « EPG prominence: A report on the discoverability of PSB and local TV services », Londres : Ofcom, 2018. Les données relatives à l'effet de mise en valeur sur le visionnage de la VOD sont moins complètes mais suggèrent une corrélation similaire.



De la même manière, on peut envisager un type d'intervention plus profond qui implique une sorte d'orientation des utilisateurs quant au contenu local, régional ou national « pertinent » et « de qualité », en s'assurant qu'ils consomment ensuite le « bon mélange »³⁶⁸. Deux questions essentielles se posent dans ce contexte – en rapport avec la prise de conscience et la fortuité – à savoir : « Les gens sont-ils conscients de toute la gamme des possibilités de contenu qui leur sont offertes en ligne, et combien de fois tombent-ils sur un contenu qu'ils aiment mais dont ils ignoraient l'existence ? »³⁶⁹ L'Ofcom britannique a montré que les obstacles à la prise de conscience et à la sérendipité peuvent être importants³⁷⁰.

L'un des moyens d'y parvenir consiste à utiliser les systèmes de radiodiffusion de service public existants³⁷¹. On peut d'abord penser à une version actualisée du GEP comme outil pour renforcer la mise en valeur de la marque des radiodiffuseurs de service public et de l'offre de contenu local, régional et national. Foster et Broughton voient ce « coup de pouce » comme un processus en deux étapes par lequel les téléspectateurs sont attirés par la chaîne ou la marque de service public, puis une série de techniques sont utilisées pour « amener les spectateurs à un éventail de contenus plus large que celui qu'ils auraient pu choisir eux-mêmes »³⁷². Les auteurs justifient la nécessité d'une nouvelle législation qui assurera la mise en valeur des marques de service public ou des marques de service individuel³⁷³ sur les plateformes en ligne. Les exigences en matière de visibilité doivent s'appliquer aux éléments essentiels de toute interface avec les consommateurs, tels qu'une grille de programmes ou un menu de services à la demande, et chaque service public de radiodiffusion doit s'attendre à obtenir au moins une icône/un bouton sur la première page d'un guide à la demande ou de son équivalent³⁷⁴. Le même raisonnement peut être appliqué aux œuvres européennes³⁷⁵.

³⁶⁸ Helberger (2011), p. 346. Justifiant également une telle approche, voir Sunstein C. R., « Television and the public interest », *California Law Review* 88. pp. 499–563.

³⁶⁹ Ofcom « Ofcom's Second Public Service Broadcasting Review, Phase Two: Preparing for the Digital Future », Londres : Ofcom, 2008, paragraphe 3.95.

³⁷⁰ Ofcom (2008), paragraphe 3.98.

³⁷¹ Pour une analyse complète, voir Burri (2015).

³⁷² Foster et Broughton *op.cit.* p. 11.

³⁷³ Foster et Broughton *op.cit.* se prononcent contre la mise en valeur de programmes individuels au motif qu'elle risque de fragmenter l'expérience de l'utilisateur et de nuire à la marque globale du radiodiffuseur de service public.

³⁷⁴ Foster et Broughton *op.cit.* p. 4.

³⁷⁵ La Directive SMA modifiée contient dans son article 13 une telle règle : « Les États membres veillent à ce que les fournisseurs de services de médias relevant de leur compétence qui fournissent des services de médias audiovisuels à la demande proposent une part d'au moins 30 % d'œuvres européennes dans leurs catalogues et mettent ces œuvres en valeur ». Le considérant 34 précise que : « ...L'identification dans les métadonnées des contenus audiovisuels qui constituent des œuvres européennes devrait être encouragée, de façon à ce que ces métadonnées soient disponibles pour les fournisseurs de services de médias. La mise en valeur suppose de promouvoir les œuvres européennes en facilitant l'accès à celles-ci. La mise en valeur peut être assurée par différents moyens, comme consacrer aux œuvres européennes une rubrique spécifique accessible depuis la page d'accueil du service, prévoir un critère de recherche « œuvres européennes » dans l'outil de recherche de ce service, utiliser des œuvres européennes dans les campagnes promotionnelles de ce service ou promouvoir un pourcentage minimal d'œuvres européennes du catalogue de ce service, par exemple à l'aide de bannières ou d'outils similaires ».



La deuxième proposition (favoriser la sérendipité) peut également être utile, « notamment pour présenter aux téléspectateurs des contenus qu'ils ne rechercheraient pas autrement, ou pour remettre en question les opinions des utilisateurs et élargir leurs connaissances "par hasard" »³⁷⁶. Dans ce contexte, un chercheur a souligné que « des rencontres fortuites pourraient apaiser certaines inquiétudes concernant les stratégies d'adaptation restrictives et la tendance des utilisateurs à se cacher dans leurs "cocons d'information" »³⁷⁷ et "promouvoir la compréhension" et l'ouverture d'esprit, et ainsi faire progresser les objectifs démocratiques »³⁷⁸. L'espace numérique et les différentes manières d'analyser les données et d'agréger le contenu permettent la diffusion aléatoire de différents types de contenu, qui peuvent être affichés à côté du contenu « choisi par le spectateur » ou dans des listes dédiées « moins recherchées », « moins regardées » et d'autres listes « moins populaires », qui ne sont pas « grand public ». En outre, comme il semble qu'il y ait une grande différence dans la disponibilité et la possibilité de découvrir des genres de contenu distincts (par exemple, les sports par rapport aux programmes éducatifs), il peut être approprié d'établir des liens entre les genres, de manière à mettre en évidence ce type de contenu et à augmenter les chances d'une consommation globale plus diversifiée³⁷⁹.

Toutefois, il convient d'être prudent en ce qui concerne ces offres aléatoires car elles peuvent être simplement ignorées, voire perturber l'expérience du téléspectateur. Des recherches ont montré que les rencontres fortuites ne doivent pas être que le fruit du hasard. Schönbach explique que pour fonctionner et motiver les utilisateurs, les surprises doivent être « ancrées dans le familier »³⁸⁰. Helberger explique en outre que « pour que l'exposition fortuite à l'information ait un sens, l'information doit correspondre à une connaissance, un intérêt ou une expérience antérieurs de l'utilisateur »³⁸¹. Dans le même ordre d'idées, Hoffman *et al.* estiment que : nous ne pouvons parler « d'expérience de la diversité » que si les utilisateurs « perçoivent et digèrent ce contenu en fonction de leurs motivations, de leur conscience et de leurs capacités »³⁸². Concevoir des outils qui fonctionnent bien à cette fin peut être une tâche difficile qui est en partie liée au thème de l'éducation aux médias. Ces outils peuvent également être liés à certaines fonctions de conception algorithmique comme « les coups de pouce d'autonomisation, qui favorisent la

³⁷⁶ Ofcom (2008a). Paragraphe 3.99–3.101.

³⁷⁷ Helberger (2011a), p. 454.

³⁷⁸ *Ibid.*, en référence à Sunstein (2007). pp. 27–28.

³⁷⁹ Pour une expérience sur la promotion de la diversité des contenus par des systèmes de recommandation, voir Möller J. *et al.* « Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity », *Information, Communication and Society*, 2018.

³⁸⁰ Schönbach K. « The own in the foreign: Reliable surprise – An important function of the Media? », *Media, Culture and Society* 29. pp. 344–353, 2007.

³⁸¹ Helberger (2011a), p. 462.

³⁸² Hoffman *et al.* (2015) affirment que pour faire l'expérience de la diversité en ligne, les utilisateurs doivent s'efforcer d'atteindre la diversité, être conscients des conditions préalables de la diversité et être en mesure de garantir l'accès à la diversité.

prise de décision dans l'intérêt des citoyens, tel que jugé par eux-mêmes, sans introduire d'autres réglementations ou incitations ou en utilisant des mesures de manipulation »³⁸³.

Dans l'ensemble, il est possible d'envisager des mécanismes qui agissent comme de « bons agrégateurs » et favorisent la visibilité, la disponibilité et la consommation de contenus locaux, nationaux et régionaux de haute qualité et fiables sur diverses plateformes. À l'ère de l'IA, on peut supposer que la conception de tels éditeurs intelligents est réalisable. La question de l'équilibre entre la vertu de l'intervention et ses éventuels effets secondaires intrinsèques à de telles actions paternalistes demeure et doit être abordée avec prudence³⁸⁴.

4.4. Observations finales

Il est évident que le paysage médiatique a déjà profondément changé et continue à évoluer. Les fonctions éditoriales des médias connaissent une évolution discrète qui est véritablement disruptive pour la production, la distribution et la consommation de médias. Les plateformes numériques, telles que Facebook et Google, ont assumé, bien que dans une mesure différente, des fonctions clés dans la médiation de contenu et ont ainsi commencé à jouer un rôle vital dans la réalisation d'objectifs publics essentiels, y compris dans le domaine culturel. Comme ils affectent la disponibilité et l'accès aux contenus locaux, nationaux et régionaux, ces intermédiaires peuvent également avoir une incidence sur la forme et le contenu des échanges culturels, sur la participation démocratique et sur l'engagement civique. Au cours des dernières années, on a pris conscience des risques liés au filtrage algorithmique et aux régimes médiatiques sur mesure qui peuvent être sévèrement restreints et/ou commercialisés. Des étiquettes telles que « bulles de filtres » ou « chambres d'écho » ont attiré l'attention des universitaires et des décideurs politiques. Toutefois, la question des effets réels de la médiation par le biais des plateformes numériques et du lien de causalité entre les différents types d'exposition aux médias et l'engagement culturel, politique et social, reste ouverte. Cela semble être le cas même pour les contenus plus simplement « mauvais », comme les fausses nouvelles. En ce sens, deux pistes de réflexion peuvent être présentées aux décideurs politiques :

- (1) nous avons besoin de plus de données et de recherches indépendantes sur la disponibilité des différents types de contenu, la consommation et l'engagement vis-à-vis de ce contenu, les participants à ce processus et l'impact de ces processus sur les performances démocratiques et culturelles individuelles et collectives ;

³⁸³ Hansen P. G. et Jespersen A. M. « Nudge and the manipulation of choice : A framework for the responsible use of the nudge approach to behaviour change in public policy », *European Journal of Risk Regulation* 1, pp. 3–28, 2013, p. 24. Pour une analyse complète de la diversité des expositions, voir Helberger N., Karppinen K., et D'Acunto L., « Exposure diversity as a design principle for recommender systems », *Information, Communication and Society*, 2017.

³⁸⁴ Helberger N., « Merely facilitating or actively stimulating diverse media choices? Public service media at the crossroad », *International Journal of Communication* 9. pp. 1324–1340, 2015 ; Bodo B. et al. « Tackling the algorithmic control crisis – The technical, legal and ethical challenges of research into algorithmic agents », *Yale Journal of Law and Technology* 19. pp. 133–180, 2017.

- (2) il est important de poursuivre le dialogue entre les créateurs de contenu, les intermédiaires, les utilisateurs, les annonceurs et les autres parties prenantes impliquées dans la dynamique de l'espace médiatique et de souligner l'importance cruciale d'une consommation de médias culturellement diversifiés dans ce dialogue. La valeur accrue attachée à la disponibilité de choix culturellement diversifiés, l'accent mis sur la fiabilité et la qualité que les utilisateurs comprennent et apprécient, pourraient très bien inciter les plateformes à offrir de telles options. Des mesures ont même déjà été prises dans ce sens - par exemple, en ce qui concerne le signalement ou le retrait de certains types de contenu ou de certains utilisateurs, ou l'amélioration de la transparence quant aux sources de contenu.

D'autres formes d'action seront néanmoins probablement nécessaires. Nous esquissons deux pistes possibles qui pourraient façonner la consommation de médias – la gouvernance *des et par le biais des* algorithmes. Pour cette dernière, on peut considérer que les « bons agrégateurs » favorisent la visibilité, la disponibilité et la consommation d'un contenu local, national et régional fiable et de haute qualité. Bien que cela puisse paraître interventionniste et comme une possible interférence avec l'autonomie des utilisateurs et la liberté d'expression, ainsi qu'avec la liberté d'entreprise des plateformes, on pourrait envisager un modèle sensible à la diversité qui ne soit pas en contradiction avec des choix autonomes mais permette véritablement aux utilisateurs de faire des choix mieux informés. La technologie est susceptible de permettre de nombreuses variations sur le thème et les décideurs politiques devront faire preuve d'ouverture d'esprit à cet égard, et pourraient envisager de confier aux médias de service public le rôle de « conservateurs » de l'expérience médiatique. La prudence reste de mise et la poursuite des objectifs de diversité ne s'inscrit pas nécessairement dans la conception pratique de tous les systèmes de recommandation – dans le cas des moteurs de recherche comme Google, par lesquels les utilisateurs recherchent activement des réponses, il doit être possible de trouver un compromis entre précision et diversité³⁸⁵.

En ce qui concerne le rôle des intermédiaires eux-mêmes et l'atténuation des risques posés par une consommation de médias sur mesure et potentiellement faussée, il sera peut-être nécessaire d'agir dans l'intérêt public. Cependant, nous ne pouvons pas rejeter toute la faute sur les plateformes ou les systèmes de recommandation et n'adopter que des mesures les ciblant. Comme le note Helberger (2017) *et al.*, les utilisateurs « jouent également un rôle dans la réalisation ou l'érosion des valeurs publiques sur ces plateformes »³⁸⁶. En effet, nous sommes confrontés à la « question de la responsabilité partagée » et il est donc nécessaire de conceptualiser un cadre avec la participation de toutes les parties prenantes (plateformes, utilisateurs, société civile et gouvernements) auxquelles sont confiées différentes responsabilités³⁸⁷. Les mécanismes multipartites

³⁸⁵ Adomavicius G. et Kwon Y., « Maximizing aggregate recommendation diversity: A graph-theoretic approach », *Proceedings of the 1st International Workshop on Novelty and Diversity in Recommender Systems*. DiveRS 2011. Chicago, 2011.

³⁸⁶ Helberger N., Pierson J. et Poell T., « Governing online platforms: From contested to cooperative responsibility », *The Information Society*, 2017, p. 2.

³⁸⁷ *Ibid.*

dérivés de la gouvernance d'internet peuvent servir de modèle³⁸⁸. L'expérience récemment acquise dans le domaine de la lutte contre la désinformation en ligne en Europe peut donner des indications particulièrement utiles. La concrétisation d'une disponibilité de contenus diversifiés et de choix d'utilisateurs informés et responsabilisés en tant que valeurs publiques fondamentales dans l'espace médiatique devrait alors être le résultat d'une interaction et d'une délibération dynamiques entre les parties prenantes et pourrait déboucher sur un éventail de mesures, telles que des codes de conduite, des lignes directrices et des principes, des organes de contrôle de nature gouvernementale ou non gouvernementale assurant un dialogue continu et efficace, ou encore certaines solutions technologiques³⁸⁹.

³⁸⁸ Voir, par exemple, Marda V. et Milan S., « Wisdom of the crowd: Multistakeholder perspective on the fake news debate », *A Report by the Internet Policy Observatory at the Annenberg School, University of Pennsylvania*, 21 mai 2018.

³⁸⁹ Helberger *et al.* (2017).

Droit d'auteur

*L'une des principales craintes que suscite l'IA est celle du remplacement des humains par les machines. Les gens ont de plus en plus peur de perdre leur emploi au profit de robots, et cette inquiétude se ressent également dans le secteur audiovisuel. Les exemples d'intervention créative de l'IA dans la rédaction de scénarios ou la composition de musique se multiplient, pour ne citer que ces deux domaines. Cette invasion techno-barbare dans le champ créatif reste toutefois d'une importance plutôt modérée, et les craintes d'une destruction des emplois créatifs sont ainsi très probablement injustifiées, du moins pour le moment. Et pourtant, le sujet de la protection par le droit d'auteur des œuvres créées par des machines enflamme le monde universitaire. Cette question n'est pas sans mérite : si l'on admet que des machines puissent « créer » des œuvres, peuvent-elles alors détenir des droits d'auteur ? Ou alors, des personnes ou des entreprises peuvent-elles être titulaires des droits d'auteur sur une œuvre créée par une machine ? L'état des lieux proposé par **Giancarlo Frosio** dans sa contribution à cette publication répond à une série de questions juridiques émergentes concernant l'IA et la créativité. L'IA peut également servir à lutter contre les ennemis de la créativité en permettant d'identifier et de retirer d'internet des contenus qui contreviennent au droit d'auteur et de traquer les pirates et les fuites provenant du secteur. Cependant, en fonction de la manière dont les algorithmes sont programmés, il existe toujours un risque d'obtenir de faux positifs, ce qui peut avoir un impact sur la liberté d'expression des utilisateurs d'internet.*

5. Droit d'auteur : la machine est-elle un auteur ?

Giancarlo Frosio, Centre d'Études Internationales de la Propriété Intellectuelle (CEIPI), Université de Strasbourg *

5.1. Introduction

D'aucuns affirment que l'intelligence artificielle (IA) est une révolution porteuse de transformations radicales pour l'humanité³⁹⁰. Les machines intelligentes prennent des formes multiples et variées pour servir divers objectifs et ont la capacité de remplacer les humains pratiquement partout³⁹¹ ce qui, bien entendu, a aussi bien des effets positifs que des effets négatifs³⁹². Il semblerait même que l'IA soit en mesure de se substituer à des activités plus foncièrement humaines. Même si jusqu'à présent, la plupart des créateurs ne se sentent pas réellement menacés par l'activité des robots³⁹³, la créativité est néanmoins un domaine majeur dans lequel l'IA s'avère de plus en plus compétente. L'IA écrit des poèmes, des romans et des articles de presse, compose de la musique, édite des photographies, crée des jeux vidéo et produit des peintures et autres œuvres d'art. La

*Maître de conférence au Centre d'Études Internationales de la Propriété Intellectuelle (CEIPI), Université de Strasbourg, chercheur non résident au Stanford Law School Center for Internet and Society, professeur associé au NEXA Center for Internet and Society. Je tiens à remercier mon assistante de recherche, Varnita Singh, pour ses recherches approfondies et son soutien exceptionnel et fondamental dans la préparation de ce chapitre.

³⁹⁰ Floridi L., *The Forth Revolution: How the Infosphere is Reshaping Human Reality*, OUP, Oxford. Bughin J. et al. (2017), *Artificial Intelligence: The Next Digital Frontier?*, McKinsey Global Institute Discussion Paper, www.mckinsey.com/-/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx. Elsevier, *Artificial Intelligence: How knowledge is created, transferred, and used trends in China, Europe, and the United States*, www.elsevier.com/research-intelligence/resource-library/ai-report. Ménière Y., Rudyk I. et Valdes J., *Patents and the Fourth Industrial Revolution: The Inventions behind Digital Transformation*, Munich, DE: Office européen des brevets.

³⁹¹ UIT, *Assessing the Economic Impact of Artificial Intelligence*, Issue Paper n° 1, Union internationale des télécommunications, Genève, pp. 12-15, www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-ISSUEPAPER-2018-1-PDF-E.pdf. Pricewaterhouse Coopers (PwC), *Sizing the prize: PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution*, PwC, www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.

³⁹² UIT, *op. cit.*, pp. 17-20.

³⁹³ Pfeiffer A., *Pfeiffer Report: Creativity and technology in the in age of AI*, pp. 15, 29, <https://www.pfeifferreport.com/wp-content/uploads/2018/10/Creativity-and-technology-in-the-age-of-AI.pdf>.



plupart des secteurs créatifs seront largement touchés³⁹⁴, depuis l'industrie audiovisuelle³⁹⁵ jusqu'à la musique³⁹⁶ et l'édition³⁹⁷. L'ère de l'(!)Auteur est en marche.

Dans ce contexte, l'évolution du système de propriété intellectuelle (PI) pour s'adapter à la créativité et à l'innovation générées par l'IA (et aux défis que cela entraîne) est désormais un thème d'une actualité cruciale³⁹⁸. On assiste actuellement à l'émergence d'une abondante littérature consacrée à l'IA et la PI³⁹⁹. Il va sans dire que les régimes de propriété intellectuelle existants, notamment en matière de droit d'auteur, de secret des affaires et de droit des brevets⁴⁰⁰, peuvent protéger les logiciels sur lesquels repose la technologie de l'IA⁴⁰¹. Cependant, la protection accordée aux logiciels ne s'étend pas aux produits éventuellement créés par l'IA. La question reste ouverte quant à savoir si, sur la base de la configuration actuelle du régime du droit d'auteur, cette protection est effectivement disponible. Il convient, par ailleurs, de distinguer la créativité assistée par ordinateur, qui est protégée par le droit d'auteur pour autant que la contribution de l'utilisateur soit originale, de la créativité proprement générée par ordinateur, où l'interaction de l'utilisateur avec un ordinateur amène ce dernier à créer sa propre

³⁹⁴ New European Media (2019), *AI in media and creative industries*, <https://arxiv.org/ftp/arxiv/papers/1905/1905.04175.pdf>. Pfeiffer A., *op.cit.*

³⁹⁵ *L'intelligence artificielle dans l'industrie audiovisuelle*, résumé de l'atelier de l'Observatoire européen de l'audiovisuel (en anglais) Strasbourg, 17 décembre 2019, Observatoire européen de l'audiovisuel, Strasbourg, 2019, <https://rm.coe.int/summary-workshop-2019-bat-2/16809c992a>. Voir aussi Baujard T., Tereszkievicz R., de Swarte A., Tuovinen T., "Entering the new paradigm of artificial intelligence and series", étude commandée par le Conseil de l'Europe et, Décembre 2019, <https://rm.coe.int/eurimages-entering-the-new-paradigm-051219/1680995331>.

³⁹⁶ Strum B. et al., *Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis*, Arts 8, pp. 115-129. BPI, *Music's smart future: How will AI Impact the Music Industry*, www.musicbank.co.uk/wp-content/uploads/2018/03/bpi-ai-report.pdf.

³⁹⁷ Lovrinovic C. et Volland H., *The future impact of artificial intelligence on the publishing industry*, Gould Finch and Frankfurter Buchmesse, disponible sur : <https://bluesyemre.files.wordpress.com/2019/11/the-future-impact-of-artificial-intelligence-on-the-publishing-industry.pdf>.

³⁹⁸ Cubert J.A. et Bone R.G.A., « The law of intellectual property created by artificial intelligence » dans Barfield W. and Pagallo U. (éd.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar, Cheltenham, pp. 411-427. OCDE, *L'intelligence artificielle dans la société*, Éditions OCDE, Paris, pp. 121-122, https://www.oecd-ilibrary.org/science-and-technology/l-intelligence-artificielle-dans-la-societe_b7f8cd16-fr. OMPI (2019a), *Projet de document de synthèse sur les politiques en matière de propriété intellectuelle et l'intelligence artificielle* OMPI, Genève, https://www.wipo.int/meetings/fr/doc_details.jsp?doc_id=470053. OMPI (2019b), *WIPO Technology Trends 2019 - Artificial Intelligence*, WIPO, Geneva, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf.

³⁹⁹ Iglesias M., Shamuilia S. et Anderberg A., *Intellectual Property and Artificial Intelligence: A Literature Review*, Office des publications de l'Union européenne, Luxembourg, https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119102/intellectual_property_and_artificial_intelligence_jrc_template_final.pdf.

⁴⁰⁰ Qu'il s'agisse de la protection des logiciels en tant qu'invention mise en œuvre par ordinateur ou en tant que tels, selon la juridiction.

⁴⁰¹ Calvin N. et Leung J., *Who owns artificial intelligence ? A preliminary analysis of corporate intellectual property strategies and why they matter*, Future of Humanity Institute, Université d'Oxford, https://www.fhi.ox.ac.uk/wp-content/uploads/Patents_-FHI-Working-Paper-Final-.pdf.

expression⁴⁰². Un rapport de la Commission européenne énonce clairement les termes de ce bourbier naissant :

La protection des œuvres générées par l'IA [...] semble être [...] problématique. À la lumière de l'approche humaniste du droit d'auteur, il est douteux que les œuvres générées par l'IA méritent une protection au titre du droit d'auteur. [...] Alors que certains spécialistes du droit d'auteur plaident clairement pour que les œuvres générées par l'IA soient placées dans le domaine public, d'autres avancent une série de propositions visant à garantir un certain niveau de protection. À quelques rares exceptions près, ces propositions [...] ne détaillent pas toujours suffisamment les éléments susceptibles de justifier cette protection⁴⁰³.

À cet égard, le présent chapitre vise à répondre à une série de questions juridiques émergentes face au dilemme de la créativité artificielle. Comment la créativité artificielle s'inscrit-elle dans la théorie conventionnelle du droit d'auteur et dans les doctrines existantes ? En particulier, quelles sont les conditions requises pour la protection des créations générées par l'IA et les réseaux de neurones profonds dans le cadre des principaux régimes de droit d'auteur ? Faut-il envisager la création d'une personne morale pour l'IA ? L'IA est-elle un auteur selon les normes traditionnelles du droit d'auteur ? Une machine peut-elle être originale ? Ces interrogations - qui ne sont que quelques-unes des nombreuses questions posées par la créativité artificielle - peuvent se résumer en une phrase : l'IA peut-elle être un (I)Auteur ? Par ailleurs, deux autres questions fondamentales dépassent le cadre de la présente analyse et concernent la machine apprenante et le contrevenant (IA). Elles visent à savoir d'une part, si une IA peut enfreindre les droits d'auteur par le biais du processus d'apprentissage automatique et d'entraînement qui lui permet de générer de la créativité, et d'autre part, si une IA peut enfreindre les droits d'auteur en produisant une contrefaçon. Outre les véritables défis liés aux normes de paternité de l'IA, le présent rapport envisage la voie à suivre au vu des options politiques basées sur les différentes approches théoriques, telles que les théories de la personnalité et les théories utilitaristes/incitatives de la propriété intellectuelle.

5.2. Technologie

Le tout premier livre écrit par ordinateur est *The Policeman's Beard is Half Constructed: Computer Prose and Poetry* de Racter⁴⁰⁴. C'était en 1984 et la prose de Racter était encore assez obscure et rudimentaire. Mais depuis, les choses ont changé. La qualité de la créativité artificielle s'est considérablement améliorée, à tel point qu'un roman écrit par une machine a franchi les premières étapes de sélection d'un concours littéraire au Japon,

⁴⁰² *Payer Components South Africa Ltd c. Bovic Gaskins* [1995] 33 IPR 407. Clark R. and Smyth S., *Intellectual Property Law in Ireland*, Butterworths, Dublin. Denicola R., « Ex Machina: Copyright Protection for Computer-Generated Works », *Rutgers University Law Review* 69, pp. 269-270.

⁴⁰³ Craglia M., *Artificial Intelligence: A European Perspective*, Office des publications de l'Union européenne, Luxembourg, pp. 67-68.

⁴⁰⁴ Racter, *The Policeman's Beard is Half Constructed: Computer Prose and Poetry* by Racter - *The First Book Ever Written by a Computer*, Warner Books, New York.

éliminant dans la foulée des milliers d'auteurs humains⁴⁰⁵, et que *Sunspring*, un film de science-fiction entièrement écrit par une IA, s'est classé dans le top 10 du festival annuel Sci-Fi London⁴⁰⁶. L'AIVA - de même que Melodrive ou Amper - gère une IA qui compose de la musique commercialisée pour accompagner des œuvres audiovisuelles, des publicités ou des jeux vidéo⁴⁰⁷. Les Z-Machines, groupe de robots musiciens japonais, interprètent de la musique en changeant le rythme de leur jeu en fonction de l'intervention du public et des personnes qui visitent leur site internet⁴⁰⁸, tandis que le Flow Machine de Sony peut interagir et co-improviser avec un musicien humain⁴⁰⁹. Les arts visuels semblent toutefois être le domaine de création où l'IA est la plus performante. Un tableau généré par IA, intitulé « Portrait d'Edmond de Belamy », a été vendu chez Christie's pour la somme impressionnante de 432 500 dollars⁴¹⁰.

Grâce à la disponibilité massive des données, à l'évolution des ressources de calcul et aux nouvelles architectures basées sur l'apprentissage profond, l'IA a connu des avancées majeures au cours de la dernière décennie⁴¹¹. L'avènement du *Generative Adversarial Network* (réseau antagoniste génératif - GAN), étroitement lié à ces progrès, a entraîné un développement fondamental de la créativité générée par l'IA⁴¹². Il s'agit là d'une évolution assez récente. En juin 2014, Ian Goodfellow a publié un article intitulé « *Generative Adversarial Networks* » et placé le code sur GitHub sous licence BSD⁴¹³. L'article décrit un processus génératif qui utilise un modèle antagoniste pour l'apprentissage automatique. Dans ce scénario, deux réseaux de neurones s'affrontent dans un jeu. Cette technique apprend à générer de nouvelles données avec les mêmes statistiques que celles de son module d'entraînement. Cette méthode est devenue très populaire pour entraîner l'IA avec de grands ensembles de données. La technologie a ensuite évolué vers des systèmes créatifs de type *Creative Adversarial Network* (réseau antagoniste créatif - CAN), qui s'appuient sur des GAN et qui « produisent de l'art en regardant l'art et en apprenant les fondements du style, puis deviennent créatifs en renforçant la charge interpellatrice de l'art produit tout en s'écartant des styles appris⁴¹⁴ ». Les GAN et les CAN ont été déployés par le collectif artistique parisien Obvious pour créer le « Portrait d'Edmond de Belamy » et une série de portraits algorithmiques intitulée « La Famille de Belamy »⁴¹⁵.

⁴⁰⁵ Lewis D., « An AI-Written Novella Almost Won a Literary Prize », *Smithsonian Magazine*, Washington, <https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577>.

⁴⁰⁶ Craig C. et Kerr I., « The Death of the AI Author », *Osgoode Legal Studies Research Paper*, pp. 1-2.

⁴⁰⁷ AIVA, <https://www.aiva.ai>.

⁴⁰⁸ Bakare L., « Meet Z-Machines, Squarepusher's new robot band », *The Guardian*, <https://www.theguardian.com/music/2014/apr/04/squarepusher-z-machines-music-for-robots>.

⁴⁰⁹ Deltorn J.M. et Macrez F. (2018), « Authorship in the Age of Machine Learning and Artificial Intelligence », *Center for International Intellectual Property Studies Research Paper* n° 2018-10, 22-23, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3261329.

⁴¹⁰ Craig C. et Kerr I., *op.cit.*, 3-4.

⁴¹¹ Goodfellow I., Bengio Y. et Courville A., *Deep Learning*, MIT Press, Cambridge.

⁴¹² Svedman M., « Artificial Creativity: A case against copyright for AI-created visual work », *IP Theory* 9(1), pp. 3-4.

⁴¹³ Goodfellow I. et al, « Generative Adversarial Networks », *arXiv*, <https://arxiv.org/abs/1406.2661>.

⁴¹⁴ Elgammal A. et al, « CAN : Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms », pp. 1-22, <https://arxiv.org/abs/1706.07068>.

⁴¹⁵ Obvious AI & Art. Disponible sur : <https://obvious-art.com>.

À l'instar de Deep Mind de Google, qui génère et interprète de la musique ou crée des œuvres d'art, l'IA le fait en écoutant d'autres musiques ou en analysant d'anciennes œuvres d'art en ligne. Pindar Van Arman entraîne depuis déjà un certain temps une IA à être créative. Ce projet, baptisé cloudpainter.com, illustre clairement les similitudes existantes entre les processus d'apprentissage humain et automatique pour créer de l'art⁴¹⁶. Il semblerait qu'une IA apprend à générer de la créativité grâce à un processus d'apprentissage en plusieurs étapes, en commençant par des exercices techniques, tels que la réalisation d'images en reliant des points, pour passer ensuite à l'expérimentation, à l'imitation et, enfin, à la création indépendante.

5.3. Protection : la créativité artificielle peut-elle être protégée ?

L'IA transforme notre façon de créer et a une incidence sur les concepts et les doctrines du droit d'auteur établis de longue date. Cela pose en particulier des questions essentielles concernant l'éligibilité de la créativité artificielle à une protection dans le cadre du régime actuel du droit d'auteur. On peut aborder ce problème spécifique au regard des trois principaux prérequis pour la protection et la titularité des droits attachés aux œuvres relevant du droit d'auteur : la personnalité juridique, la paternité et l'originalité.

5.3.1. Personnalité : une machine peut-elle être une personne morale ?

La première question pertinente consiste à savoir si les machines peuvent jouir d'une personnalité juridique. Selon la juridiction, les valeurs culturelles, les croyances religieuses et la subjectivité juridique, la reconnaissance de la personnalité des machines peut relever d'un choix politique. Le Japon a toujours entretenu une relation particulière avec les robots et les machines en raison des croyances shintoïstes selon lesquelles on peut imaginer que les robots imitant des animaux ou des humains ont une âme⁴¹⁷. En octobre 2017, Sophia est devenue le premier robot à se voir accorder la citoyenneté par le gouvernement saoudien. Cette mesure était manifestement un « coup de pub » destiné à faire sensation. Néanmoins, il s'agit d'une étape historique vers une possible intégration de l'IA à l'humanité. Et, un mois plus tard, en novembre 2017, Tokyo a accordé à un chatbot le statut de résident officiel dans le quartier de Shibuya⁴¹⁸.

⁴¹⁶ Cloud Painter. Disponible sur : <https://www.cloudpainter.com>.

⁴¹⁷ Holland-Minkley D., *God in the Machine: Perceptions and Portrayals of Mechanical Kami in Japanese Anime*, mémoire de maîtrise, Université de Pittsburgh.

⁴¹⁸ Cuthbertson, « Tokyo : Artificial Intelligence 'Boy' Shibuya Mirai Becomes World's First AI Bot to Be Granted Residency », *Newsweek*, Washington, <https://www.newsweek.com/tokyo-residency-artificial-intelligence-boy-shibuya-mirai-702382>.

L'idée d'une personnalité juridique propre aux machines intelligentes est également étayée par une réflexion théorique. Nick Bostrom, par exemple, note que : « les machines capables de prendre des initiatives autonomes et d'élaborer leurs propres plans ... sont davantage considérées comme des personnes que comme des machines⁴¹⁹. » Certains auteurs soulignent qu'il n'existe aucun fondement juridique ou théorique pour réfuter la personnalité des robots IA : la loi devrait pouvoir accorder la personnalité sur la base de choix rationnels et de données empiriques, plutôt que sur la base de superstitions et de privilèges⁴²⁰. Par conséquent, des arguments ont été avancés en faveur de l'octroi d'une personnalité à de futures IA potentiellement performantes qui seraient autonomes (capables de prendre une décision sans intervention extérieure), intelligentes (capables de s'auto-programmer et d'intégrer des informations dans un cadre spécifique) et qui auraient une conscience (capables d'une expérience subjective)⁴²¹.

Fait notable, le Parlement européen envisage la possibilité de qualifier l'IA et les robots comme des « personnes électroniques ». Dans une résolution sur les règles de droit civil en matière de robotique, le Parlement européen se demande si les règles ordinaires en matière de responsabilité sont suffisantes ou si de nouveaux principes et de nouvelles règles s'imposent pour encadrer l'IA⁴²². La nature autonome des robots doit-elle être interprétée à la lumière des catégories juridiques existantes ou faut-il créer une nouvelle catégorie⁴²³ ? La résolution affirme que « plus un robot est autonome, moins il peut être considéré comme un simple outil contrôlé par un autre acteur (le fabricant, l'opérateur, le propriétaire, l'utilisateur, etc.)⁴²⁴ ». Aux yeux du Parlement européen, il est évident que le droit de l'UE ne permet pas de couvrir la responsabilité non contractuelle en cas de dommages causés par une IA autonome. Les règles traditionnelles suffisent à couvrir les cas où la cause des actes ou de l'inaction du robot peut être identifiée comme imputable à un acteur humain précis, tel que le fabricant, l'opérateur, le propriétaire ou l'utilisateur. De même, les règles actuellement en vigueur en matière de responsabilité s'appliquent aussi en cas de dysfonctionnement du robot ou si l'agent humain aurait pu prévoir et éviter le comportement dommageable du robot. Mais que se passe-t-il si la cause de l'acte ou de l'inaction du robot ne peut être attribuée à un agent humain spécifique ? Qu'en est-il lorsqu'il n'y a aucun défaut de fabrication ? Ou qu'il n'y a pas eu de dysfonctionnement de l'IA ? Que faire lorsque la victime n'est pas en mesure d'apporter des preuves des dommages effectifs, du défaut du produit ou de la relation de cause à effet entre les dommages et le défaut ? Que peut-on faire si, en fait, l'IA a causé des dommages parce qu'elle a agi de manière autonome selon sa propre programmation et son propre objectif ? Dans un tel

⁴¹⁹ Bostrom N., *Superintelligence: Paths, Dangers, Strategies*, OUP, Oxford.

⁴²⁰ Solum L.B., « Legal Personhood for Artificial Intelligences », *North Carolina Law Review* 70, p. 1264.

⁴²¹ Zimmerman E., « Machine Minds : Frontiers in Legal Personhood », pp. 14-21, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965. Voir aussi Hubbard F.P., « Do Androids Dream? Personhood and Intelligent Artifacts », *Temple law Review* 83, pp. 406-474.

⁴²² Cf. Vladeck D., « Machines sans principes: Liability Rules and Artificial Intelligence », *Washington Law Review* 89, pp. 117-150.

⁴²³ Parlement européen (2017), Règles de droit civil sur la robotique : Résolution du Parlement européen du 16 février 2017 contenant des recommandations à la Commission concernant des règles de droit civil sur la robotique (2015/2103(INL)), 16 février 2017, https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_FR.pdf.

⁴²⁴ Parlement européen (2017), *op.cit.*

scénario, la Directive 85/374/CEE sur la responsabilité du fait des produits défectueux ne peut pas s'appliquer. La résolution souligne que de ce fait, les règles ordinaires sur la responsabilité sont insuffisantes et demande de nouvelles règles permettant de clarifier si une machine peut être tenue responsable de ses actes ou de ses inactions⁴²⁵. Tout en reconnaissant que, « du moins en l'état actuel des choses, la responsabilité doit être imputable à un humain et non au robot », à terme, la résolution demande ce qui suit :

- 1) un régime d'assurance obligatoire couvrant toutes les responsabilités potentielles sur l'ensemble de la chaîne⁴²⁶ ;
- 2) la création 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 de réparer tout dommage causé à un tiers⁴²⁷. »

La notion de personnalité juridique propre à l'IA a émergé au cours des nombreux échanges sur ce thème, mais auparavant, le débat était dominé par des ébauches de mesures incohérentes et décousues visant à réglementer une technologie dont le développement est totalement imprévisible. Comme c'est souvent le cas dans une telle situation, le discours sur l'octroi d'une personnalité juridique est devenu une question politique dénuée de tout fondement rationnel. À cet égard, l'octroi de la citoyenneté à Sophia par l'Arabie saoudite fait écho à l'élévation par l'empereur romain Caligula de son cheval favori, Incitatus, au rang de sénateur⁴²⁸.

Que l'IA qui se développe soit quasi-humaine ou hyper-humaine, la personnalité juridique des machines est clairement absente du cadre juridique actuel et il est plus que probable qu'elle le restera dans un avenir proche. Les universitaires n'ont cessé de répéter que toute hypothèse d'octroi d'une personnalité juridique aux robots IA devait être écartée jusqu'à l'émergence de changements technologiques fondamentaux⁴²⁹. Pagallo fait état, parmi les arguments normatifs contre la personnalité juridique, du « problème de carence » (*missing something problem*), selon lequel la majeure partie des conditions préalables qui sont généralement associées à l'octroi d'une personnalité juridique à quelqu'un ou quelque chose fait totalement défaut aux robots IA actuels, puisque ces agents artificiels ne sont pas conscients d'eux-mêmes, ne sont pas animés par des intentions à caractère humain et ne sont pas en mesure de souffrir à proprement parler⁴³⁰. L'analyse statistique des différentes conditions de la personnalité juridique établies par la jurisprudence américaine, par exemple, montre également une incompatibilité entre la personnalité juridique et les

⁴²⁵ Parlement européen (2017), *op.cit.* Voir également Commission européenne, Liability for Artificial Intelligence and other Emerging Technologies: Report from the Expert Group on Liability and New Technologies – New Technologies Formation, Union européenne, Bruxelles, <https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608>.

⁴²⁶ Parlement européen (2017), *op.cit.*

⁴²⁷ Parlement européen (2017), *op.cit.*

⁴²⁸ Pagallo U., « Vital, Sophia, and Co - The Quest for the Legal Personhood of Robots », *Information* 9(9), pp. 239-240.

⁴²⁹ Banteka N., « Artificially Intelligent Persons », *Houston Law Review* 58 (à paraître), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3552269. Mik E., « AI as Legal Person? » dans Hilty R. et Liu K-C. (éd.), *Artificial Intelligence and Intellectual Property*, Oxford University Press, Angleterre, à paraître. Pagallo U., *op.cit.* pp. 230-240.

⁴³⁰ Pagallo U., *op.cit.* pp. 237-238.

entités d'IA⁴³¹. Cette analyse empirique prouve que, pour accorder la personnalité, les tribunaux examinent si elle est garantie directement ou indirectement par une loi, si l'entité artificielle peut poursuivre et être poursuivie en justice, et enfin, si l'entité est un groupement de personnes physiques⁴³².

Ces considérations permettent également de faire la distinction entre l'IA et les sociétés ayant le statut de personne morale. Contrairement aux sociétés, les entités d'IA ne sont ni des entités « fictives », ni des associations de personnes physiques⁴³³. Les personnes morales sont constituées par des personnes physiques qui sont fondamentalement en mesure d'exploiter des droits. En outre, même si des personnes morales peuvent détenir un droit d'auteur, ce droit provient d'une œuvre créée par une personne physique qui est l'auteur de cette œuvre, ladite œuvre remplissant alors les critères de paternité et d'originalité. Or, il en va tout autrement avec une IA et une œuvre créée par une IA, comme nous le verrons dans les pages suivantes.

Il est probable que les décideurs politiques ont d'ores et déjà intégré les arguments s'opposant à la personnalité juridique de l'IA, comme en témoigne la résolution du Parlement européen de 2017 en excluant, ne serait-ce qu'à court et moyen terme, toute forme de personnalité juridique de l'IA. En outre, le Parlement européen semble désormais exclure la personnalité juridique de l'IA en ce qui concerne particulièrement la créativité artificielle. Dans un récent « Projet de rapport sur les droits de propriété intellectuelle pour le développement des technologies de l'intelligence artificielle », le Parlement européen note, dans le cadre d'une proposition de résolution parlementaire, que « l'autonomisation du processus créatif soulève des questions relatives à la titularité des DPI, [mais] considère à cet égard qu'il ne serait pas opportun de vouloir doter les technologies d'IA de la personnalité juridique⁴³⁴. » Plutôt que d'établir la personnalité juridique des machines, le défi politique consisterait à répartir correctement les responsabilités et obligations de rendre compte des activités des robots IA lorsque cette répartition s'avère complexe, par exemple en s'appuyant sur des contrats et sur le droit des affaires⁴³⁵.

5.3.2. Paternité : une machine peut-elle être un auteur ?

Nonobstant le fait qu'en l'absence de personnalité juridique l'IA ne puisse pas être investie de la qualité d'auteur ni être en mesure de faire valoir des droits sur la créativité qu'elle peut générer, il semble pertinent d'examiner si cette créativité est éligible à une protection dans le cadre juridique actuel. Pour répondre à la vaste question de savoir si la créativité artificielle est éligible à une protection au titre du droit d'auteur, il convient de se demander si l'IA peut être considérée comme un auteur selon les normes traditionnelles du droit

⁴³¹ Banteka N., *op. cit.*

⁴³² Banteka N., *op. cit.*, pp. 50-52.

⁴³³ Banteka N., *op. cit.* p. 19.

⁴³⁴ Parlement européen (2020), *Projet de rapport sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle*, 2020/2015(INI), 24 avril 2020, https://europarl.europa.eu/doceo/document/JURI-PR-650527_FR.html.

⁴³⁵ Parlement européen (2017), *op.cit.* Pagallo U., *op.cit.* pp. 239-240.

d'auteur. Cela se résume à la question de savoir si l'existence d'un être humain est une condition intrinsèque du statut d'auteur. Un auteur peut-il être une machine ou doit-il impérativement être humain ?

En fait, il n'existe aucune définition dans les traités internationaux qui puisse apporter une réponse définitive. Toutefois, il semblerait que le libellé de la définition de la création humaine dans la Convention de Berne⁴³⁶ exclut la possibilité de considérer l'IA comme un auteur. D'une part, le terme « protection », lié à la vie de l'auteur, semble exclure les machines du statut d'auteur (Convention de Berne, article 7). De plus, la référence à la nationalité - ou à la résidence - de l'auteur semble impliquer que la notion de paternité ne s'applique qu'aux agents humains (Convention de Berne, article 3). D'une façon générale, d'aucuns affirment que le « modèle humaniste de Berne », ancré dans les théories idéelles de la personnalité, défend résolument une « conception anthropocentrique de la paternité telle qu'elle est inscrite dans la Convention de Berne », qui exclut toute paternité extra-humaine du champ d'application de Berne⁴³⁷.

5.3.2.1. Union européenne

Si l'on procède à un examen approfondi de la législation européenne, on parvient à des conclusions similaires⁴³⁸. Il n'existe pas de définition transversale de la notion d'auteur dans le droit statuaire, néanmoins l'auteur est défini comme une personne physique, un groupe de personnes physiques ou une personne morale tant dans l'article 2, paragraphe 1 de la Directive concernant la protection juridique des programmes d'ordinateur⁴³⁹ que dans l'article 4, paragraphe 1 de la Directive relative à la protection juridique des bases de données⁴⁴⁰. En outre, l'article 2, paragraphe 1 de la Directive sur la durée de la protection du droit d'auteur⁴⁴¹ dispose que le réalisateur principal d'une œuvre cinématographique ou audiovisuelle est considéré comme l'auteur ou un des auteurs. En fait, les *travaux préparatoires* des directives concernant la protection des programmes d'ordinateur et la protection des bases de données se sont appuyés de façon plus directe sur une vision anthropocentrique de la paternité en faisant spécifiquement référence à « l'auteur humain

⁴³⁶ Convention de Berne pour la protection des œuvres littéraires et artistiques (modifiée le 28 septembre 1979), <https://wipolex.wipo.int/fr/treaties/textdetails/12214>.

⁴³⁷ Ginsburg J., « People Not Machines: Authorship and What It Mean in the Berne Convention », *International Review of Intellectual Property and Competition Law* 49, pp.131-135. Voir également Aplin T. et Pasqualetto G., « Artificial Intelligence and Copyright Protection », paragraphe 5.04, in Ballardini R., Kuoppamäki P. et Pitkänen O.(éd.), *Regulating Industrial Internet Through IPR, Data Protection and Competition Law*, Kluwer, Alphen aan den Rijn. Ricketson S., « The Need for Human Authorship - Australian Developments: Telstra Corp Ltd v Phone Directories Co Pty Ltd (case comment) », *E.I.P.R.* 34(1), p. 34. Ricketson S. (1991), « People or machines? The Berne Convention and the changing concept of authorship », *Columbia VLA Journal of Law & the Arts* 16, p. 34.

⁴³⁸ Deltorn J.M. (2017), « Deep Creations: Intellectual Property and the Automata », *Frontiers in Digital Humanities*, p. 8, <https://www.frontiersin.org/articles/10.3389/fdigh.2017.00003/full>. Deltorn J.M. et Macrez F. (2018), *op.cit.*, p. 8.

⁴³⁹ Directive 2009/24/CE du Parlement européen et du Conseil du 23 avril 2009 concernant la protection juridique des programmes d'ordinateur, JO L111/16.

⁴⁴⁰ Directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données JO L077/20.

⁴⁴¹ Directive 2006/116/CE du Parlement européen et du Conseil du 12 décembre 2006 relative à la durée de protection du droit d'auteur et de certains droits voisins, J.O. L372/12.

qui crée l'œuvre » et à « la personne physique [qui] conservera tout au moins le droit inaliénable de revendiquer la paternité de son œuvre⁴⁴². » La proposition initiale de directive relative à la protection des programmes d'ordinateur concluait ainsi : « [l']apport humain en ce qui concerne la création de programmes générés par des machines peut être relativement modeste et le sera de plus en plus à l'avenir. Néanmoins, un "auteur" humain au sens le plus large est toujours présent et doit avoir le droit de revendiquer la paternité du programme⁴⁴³. » Dans l'affaire jugée par la CJUE, *Painer*, l'avocate générale Verica Trstenjak a également souligné ce point en ces termes : « C'est donc uniquement le résultat d'une activité créatrice humaine qui est protégé, activité créatrice qui n'est pas exclue du seul fait que l'auteur se serve d'un moyen technique tel qu'un appareil photo⁴⁴⁴. »

La législation nationale des États membres de l'UE confirme cette approche. Par exemple, l'article L.111-1 du Code français de la propriété intellectuelle⁴⁴⁵ exige que l'œuvre protégée soit une « création de l'esprit ». L'article 5 de la loi espagnole sur le droit d'auteur énonce clairement que « l'auteur d'une œuvre est la personne physique qui l'a créée »⁴⁴⁶. Bien que l'article 7 de la loi allemande sur le droit d'auteur ne limite pas spécifiquement la paternité d'une œuvre aux personnes physiques, l'article 11 ancre cette paternité dans une approche fondée sur la personnalité en protégeant « l'auteur dans ses relations intellectuelles et personnelles avec l'œuvre »⁴⁴⁷.

En outre, le droit de l'UE - ainsi que de nombreuses législations nationales (par exemple, la loi néerlandaise sur le droit d'auteur, art. 4, paragraphe 1, le Code français de la propriété intellectuelle, article L113-1, la loi espagnole sur le droit d'auteur, art. 6.1, la loi italienne sur le droit d'auteur, art. 8)⁴⁴⁸ - adopte une approche anthropocentrique en prévoyant une présomption de paternité pour la *personne* dont le nom est indiqué dans l'œuvre, en l'absence de preuve du contraire (Directive sur l'application des droits de propriété intellectuelle, article 5). Certains tribunaux nationaux ont même établi que cette présomption n'était applicable qu'aux personnes physiques qui ont créé l'œuvre, et non à

⁴⁴² Ramalho A., « Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems », *Journal of Internet Law* 21, pp. 17-18.

⁴⁴³ Commission européenne, *Exposé des motifs de la proposition de directive sur la protection des programmes d'ordinateur*, COM (88) 816 final, 17 mars 1989, p. 21.

⁴⁴⁴ Avis de l'avocat général Trstenjak (12 avril 2011), C-145/10 *Eva-Maria Painer c. Standard VerlagsGmbH*, ECLI:EU:C:2011:239.

⁴⁴⁵ Code de la propriété intellectuelle de 1912. Disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/fr/fr467en.pdf> (France).

⁴⁴⁶ Loi sur la propriété intellectuelle de 1996. Disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/es/es177en.pdf> (Espagne).

⁴⁴⁷ Urheberrechtsgesetz (loi sur le droit d'auteur et les droits voisins - UrhG) 1965. Disponible sur : https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html (Allemagne).

⁴⁴⁸ Auteurswet (loi sur le droit d'auteur) 1912, disponible sur : <https://wetten.overheid.nl/BWBR0001886/2012-01-01> (Pays-Bas) ; Code de la propriété intellectuelle de 1912, disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/fr/fr467en.pdf> (France) ; loi de 1996 sur la propriété intellectuelle, disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/es/es177en.pdf> (Espagne) ; loi sur la protection du droit d'auteur et des droits voisins de 1941, disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/it/it211en.pdf> (Italie).

une personne morale qui aurait pu obtenir les droits patrimoniaux⁴⁴⁹. Théoriquement, cette présomption de paternité pourrait s'appliquer aux œuvres générées par l'IA en considérant que la ou les personnes dont le nom est indiqué dans l'œuvre en sont l'auteur ou les auteurs. Bien entendu, cette solution n'offre aucune protection réelle contre les infractions, puisque la présomption peut être réfutée en prouvant que ce n'est pas la personne désignée qui est l'auteur, mais une IA.

Un rapide tour d'horizon des autres grandes juridictions nous mène à des conclusions similaires concernant l'application de la notion de paternité à l'IA.

5.3.2.2. Australie

La loi australienne fixe un cadre statutaire très clair pour les auteurs non humains en définissant l'auteur comme une « personne qualifiée », aux termes de l'article 32, paragraphe 1 de la loi australienne sur le droit d'auteur, ce qui, en vertu de l'article 32, paragraphe 4, désigne un citoyen australien /une personne résidant en Australie⁴⁵⁰. Par ailleurs, les tribunaux australiens inscrivent l'originalité parmi les conditions préalables à la protection de la paternité. Dans l'affaire *Acohs c. Ucorp*, qui concerne l'application du droit d'auteur à des fiches techniques produites électroniquement, le tribunal spécifie qu'une œuvre doit « résulter des efforts originaux d'un auteur humain unique »⁴⁵¹. La décision rendue dans l'affaire des annuaires téléphoniques renforce ce point en soulignant que le droit d'auteur « ne s'applique que s'il émane d'un individu »⁴⁵². Enfin, dans l'affaire *IceTV c. Nine Network Australia*, qui portait sur le droit d'auteur attaché à la création par ordinateur de programmes télévisés hebdomadaires, le tribunal a conclu que seuls les auteurs, c'est-à-dire les personnes correspondant à la définition légale, pouvaient faire preuve d'originalité⁴⁵³.

5.3.2.3. États-Unis

Aux États-Unis, la question de la protection des produits de la créativité informatique n'est pas une nouveauté. Les chercheurs ont commencé à discuter de la possibilité de protéger la créativité générée par ordinateur dès la fin des années 1960⁴⁵⁴. Le Congrès américain a créé un comité chargé de déterminer si les ordinateurs ou les programmes informatiques

⁴⁴⁹ *Herlitz PBS AG c. Realister OÜ*, Cour suprême estonienne (7 février 2012) affaire n° 3-2-1-155-11 (Estonie). Voir également Vasamae E., « Presumption of authorship : only natural persons », Kluwer Copyright Blog, Amsterdam, disponible sur : http://copyrightblog.kluweriplaw.com/2012/03/19/presumption-of-authorship-only-natural-persons/?doing_wp_cron=1594514535.1866068840026855468750.

⁴⁵⁰ Copyright Act 1968, disponible sur : <https://www.legislation.gov.au/Details/C2019C00042> (Australie).

⁴⁵¹ *Acohs Pty Ltd c. Ucorp Pty Ltd* (2010) 86 IPR 492 (AUS).

⁴⁵² *Phone Directories Co Pty c. Telstra Corporation Ltd* (2010) 194 FCR 142 (AUS) Voir aussi McCutcheon J., « The Vanishing Author in Computer-Generated Works » : A Critical Analysis of Recent Australian Case Law », *Melbourne University Law Review* 36(3), pp. 941-969, https://www.researchgate.net/publication/289409001_The_vanishing_author_in_computer-generated_works_A_critical_analysis_of_recent_Australian_case_law.

⁴⁵³ *IceTV Pty Ltd c. Nine Network Australia Pty Ltd* [2009] HCA 14, 239 CLR 458 (AUS).

⁴⁵⁴ Milde K.F., « Can a Computer Be an "Author" or an "Inventor" ? », *Journal of the Patent Office Society* 51, pp. 378-406.



pouvaient être des auteurs dont la production était éligible à une protection au titre du droit d'auteur. En 1978, la *National Commission on New Technological Uses of Copyrighted Works* (Commission nationale sur les nouvelles utilisations technologiques des œuvres protégées par le droit d'auteur - Commission CONTU) a déclaré que les ordinateurs n'étaient que des « outils de création inertes qui ne produisaient pas encore d'œuvres de manière autonome ». La Commission CONTU n'a pas discuté de la protection au titre du droit d'auteur des œuvres automatisées dépourvues de paternité humaine, car à l'époque, cela relevait de la pure spéculation⁴⁵⁵. En 1986, le *Congress Office of Technology Assessment* (Bureau d'évaluation technologique du Congrès - OTA) a publié un rapport affirmant que même si les ordinateurs sont davantage que des « outils de création inertes », il est impossible de déterminer les droits d'auteur relatifs aux œuvres créées par ordinateur⁴⁵⁶.

La loi américaine sur le droit d'auteur ne comporte pas de définition juridique formelle de la paternité, de sorte que, dans un premier temps, les auteurs ont fait valoir que la loi ne limitait pas littéralement la paternité aux auteurs humains⁴⁵⁷. Néanmoins, tant les références doctrinales complémentaires que la jurisprudence excluent manifestement la possibilité de considérer les agents non humains comme des auteurs au sens de la loi. En particulier, l'article 101 de la loi sur le droit d'auteur⁴⁵⁸ définit les œuvres anonymes comme étant « celles dont la paternité n'est imputable à aucune personne physique », désignant ainsi les personnes physiques comme auteurs potentiels. De plus, il est clair que dans l'histoire constitutionnelle, le terme « *copyright* » privilégie les humains en tant que seuls « auteurs » possibles⁴⁵⁹. Les tribunaux américains ont toujours soutenu cette interprétation. La Cour suprême établit clairement qu'« en règle générale, l'auteur est [...] la *personne* qui traduit une idée en une expression fixe et tangible relevant de la protection du droit d'auteur »⁴⁶⁰. Dans l'affaire *Feist c. Rural*, la Cour suprême des États-Unis analyse en détail la notion de paternité et d'auteur en examinant le concept d'originalité qui, selon elle, se réfère à des spécificités intrinsèquement humaines, telles que « l'étincelle créative » ou « la production intellectuelle, la pensée et la conception »⁴⁶¹. Les affaires plus anciennes mènent à la même conclusion. L'examen des affaires relatives aux marques commerciales révèle que la loi sur le droit d'auteur ne protège que « les fruits du travail intellectuel » qui

⁴⁵⁵ *National Commission on New Technological Uses of Copyrighted Works* (CONTU), *Final Report*, États-Unis, p.44. Voir aussi Bridy A. (2012), « Coding Creativity: Copyright and the Artificially Intelligent Author », *Stanford Technology Law Review* 5, 22-24, paragraphe 53-60. Miller A.R., « Copyright Protection for Computer Programs, Databases, and Computer Generated Works: Is Anything New Since CONTU ? », *Harvard Law Review* 977, pp. 977-1072.

⁴⁵⁶ *US Office of Technology Assessment, Intellectual Property Rights in an Age of Electronics and Information*, États-Unis.

⁴⁵⁷ Bridy A. (2012), *op.cit.*, 20, paragraphe 49. Denicola R. (2016), *op.cit.*, pp. 275-283. Miller A.R., *op.cit.* pp. 1042-1072, Samuelson P. (1986), « Allocating Ownership Rights in Computer-Generated Works », *University of Pittsburg Law Review* 47(4), pp. 1200-1204.

⁴⁵⁸ Copyright Act de 1976. Disponible sur : <https://www.copyright.gov/title17/title17.pdf> (US).

⁴⁵⁹ Butler T., « Can a computer be an author? Copyright aspects of artificial intelligence », (Comm/Ent), *A Journal of Communications and Entertainment Law* 4(4), pp. 733-734. Clifford, R.D. (1996), « Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand up », *Tulane Law Review* 71, 1682-1686. Kasap A., « Copyright and Creative Artificial Intelligence (AI) Systems: A Twenty-First Century Approach to Authorship of AI-Generated Works in the United States », *Wake Forest Intellectual Property Law Journal* 19(4), p. 358, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3597792. Milde K.F., *op.cit.* pp. 391-392.

⁴⁶⁰ *Community for Creative Non-Violence c. Reid*, 490 U.S. 730 (1989).

⁴⁶¹ *Feist Publications c. Rural Telephone Service*, 499 U.S. 340 (1991) (États-Unis).



« sont fondés sur la force créative de l'esprit »⁴⁶². Dans l'affaire *Burrow-Giles*, la Cour suprême des États-Unis rappelle que le droit d'auteur se limite à la « conception intellectuelle originale de l'auteur »⁴⁶³.

Une affaire récente a permis d'apporter quelques éclaircissements supplémentaires sur la question. En l'occurrence, c'est un singe macaque, Naruto, qui a contribué à clarifier la situation. Dans l'affaire *Naruto contre Slater*, deux « selfies de singes » mondialement connus ont donné lieu à une polémique sur la question de savoir si les animaux pouvaient être titulaires de droits d'auteur. Les autoportraits ont été réalisés par le macaque huppé « Naruto », âgé de sept ans, alors que le photographe David Slater avait laissé son appareil photo sans surveillance lors d'un séjour en Indonésie. Peu de temps après, Wikimedia Commons a publié les photos sur son site internet en partant du principe que les selfies du singe n'avaient pas d'auteur humain et, partant, appartenaient au domaine public. Wikimedia a systématiquement refusé de retirer les photos. Revenant sur ses déclarations initiales selon lesquelles les selfies avaient été réalisés intégralement par les singes sans aucune intervention humaine, David Slater a expliqué par la suite qu'il avait au préalable effectué quelques réglages sur l'appareil pour sélectionner le bon angle et l'éclairage approprié, qu'il avait optimisé les paramètres et fait en sorte d'attirer les singes pour qu'ils appuient sur le bouton de l'appareil photo.

Même si en dernier ressort, la question de savoir si les selfies appartiennent au domaine public n'a pas été examinée, les tribunaux ont pu, à cette occasion, déterminer si Naruto était en mesure de détenir un droit d'auteur pour son selfie. En 2014, les selfies du singe ont été publiés dans un livre par Blurb Inc., qui a désigné Slater et Wildlife Personalities Ltd comme étant les titulaires des droits d'auteur. En 2015, l'association People for the Ethical Treatment of Animals (pour une éthique dans le traitement des animaux - PETA) a déposé une plainte pour violation des droits d'auteur en qualité de proche et de représentant de Naruto contre Slater, Wildlife Personalities Ltd et Blurb Inc. devant le tribunal d'instance de Californie. Le tribunal d'instance a fait droit à la demande de rejet déposée par les défenderesses au motif que Naruto n'avait pas qualité à agir en vertu de la loi sur le droit d'auteur, tout en précisant : « Si les humains qui prétendent agir au nom de la demanderesse souhaitent que le droit d'auteur fasse partie des domaines du droit où les animaux non humains ont qualité pour agir, ils devraient soumettre cette affaire contestable au Congrès - et non aux tribunaux fédéraux⁴⁶⁴. » La décision a fait l'objet d'un appel et, bien que les parties aient convenu d'un accord, la cour d'appel a refusé de rejeter l'appel et a confirmé la décision de la juridiction inférieure. La majorité des juges a constaté que si les animaux disposent de l'art. III pour leur défense, ils n'ont pas qualité à agir en vertu de la loi sur le droit d'auteur⁴⁶⁵. Le tribunal s'est appuyé sur la décision du neuvième circuit dans l'affaire *Cetacean Community. c. Bush*, dont il ressort que les animaux n'ont un statut légal que si la loi le stipule clairement⁴⁶⁶. En outre, les termes « enfants », « petits-enfants », « légitime », « veuve » et « veuf » utilisés dans la loi impliquent nécessairement que la loi sur le droit d'auteur exclut les animaux qui « ne se marient pas et n'ont pas

⁴⁶² *Trade-Mark Cases*, 100 U.S. 82 (1879) (États-Unis).

⁴⁶³ *Burrow-Giles Lithographic Co. c. Sarony*, 111 U.S. 53 (1884) (États-Unis).

⁴⁶⁴ *Naruto c. David Slater*, 15-cv-04324-WHO (N.D. Cal. 2016) (États-Unis) (« Naruto 2016 »).

⁴⁶⁵ *Naruto c. David Slater*, F.3d 418 (9^e cir. 2018) (États-Unis) (« Naruto 2018 »).

⁴⁶⁶ *Cetacean Community. c. Bush*, 386 F.3d 1169 (9^e cir. 2004) (États-Unis).

d'héritiers ayant capacité à recevoir des biens en vertu de la loi »⁴⁶⁷. Les conclusions de l'arrêt *Naruto* peuvent facilement être étendues à toute créativité non humaine et générée par l'IA.

La «*Third Edition of the Compendium of U.S. Copyright Office Practices* » (troisième édition du recueil des pratiques du Bureau du droit d'auteur des États-Unis), qui a été publiée en décembre 2014, après le début de l'affaire *Naruto*, formule à titre consultatif des recommandations d'experts qui excluent la paternité non humaine⁴⁶⁸. Le recueil fait systématiquement référence à des personnes ou à des êtres humains lorsqu'il est question de la paternité des œuvres. En particulier, l'article 306, intitulé *The Human Authorship Requirement* (Exigences en matière de paternité humaine), limite l'enregistrement des demandes aux « conceptions intellectuelles originales de l'auteur » créées par un être humain. Comme le précise l'article 313.2, *Works that Lack Human Authorship* (Œuvres dont l'auteur n'est pas un être humain), les œuvres produites par la nature, les animaux ou les plantes et, de même, les œuvres créées par une machine ou par un procédé mécanique sans l'intervention d'un auteur humain, ne sont pas protégées par le droit d'auteur. En se référant aux affaires des marques commerciales et à *Burrow-Giles*, le Copyright Office a conclu qu'il refuserait d'enregistrer une demande dès lors qu'il est établi qu'aucun être humain n'a créé l'œuvre⁴⁶⁹.

5.3.2.4. La Chine

En Chine, les questions de la paternité et de la protection au regard du droit d'auteur des œuvres créées par IA ont été débattues par de nombreux tribunaux. Le positionnement chinois concernant la paternité de l'IA s'inscrit à première vue dans la même optique que les autres juridictions, tout en laissant une certaine marge de manœuvre pour aménager une éventuelle protection. Dans l'affaire *Beijing Feilin Law Firm c. Baidu Corporation*, le tribunal a confirmé la nécessité d'un auteur humain et réfuté toute protection au titre du droit d'auteur pour les œuvres créées uniquement par des machines⁴⁷⁰. L'affaire concernait un rapport publié par la demanderesse - un cabinet d'avocats basé à Pékin - sur son compte officiel WeChat. Après qu'un internaute non identifié a publié le rapport en ligne sans autorisation, la demanderesse a déposé plainte pour contrefaçon devant le tribunal de Beijing chargé des affaires en ligne. Le rapport avait été généré à l'aide de *Wolters Kluwer China Law & Reference*, un logiciel de recherche d'informations juridiques. Alors que la demanderesse faisait valoir que l'outil ne servait qu'à des fins d'assistance, la défenderesse affirmait que l'intégralité du rapport avait été créée par ce logiciel. Le tribunal a donné gain

⁴⁶⁷ *Naruto c. David Slater*, F.3d 418 (9^e cir. 2018) (États-Unis) (« *Naruto 2018* »).

⁴⁶⁸ U.S. Copyright Office, *Compendium of U.S. Copyright Office Practices*, 3^e édition, <https://www.copyright.gov/comp3/comp-index.html>.

⁴⁶⁹ U.S. Copyright Office, *op.cit.*, paragraphe 313.2.

⁴⁷⁰ *Beijing Feilin Law Firm c. Baidu Corporation* (26 avril 2019), Tribunal de Beijing chargé des affaires en ligne, (2018) Beijing 0491 Minchu n° 239. He K. (2020b), « *Feilin c. Baidu: Beijing Internet Court tackles protection of AI/software-generated work and holds that copyright only vests in works by human authors* », *The IPKat*, <http://ipkitten.blogspot.com/2019/11/feilin-v-baidu-beijing-internet-court.html>.
Chen M., « *Beijing Internet Court denies copyright to works created solely by artificial intelligence* », *Journal of Intellectual Property Law & Practice* 14(8), pp. 14-18.

de cause à la demanderesse. Néanmoins, tout en établissant que le rapport litigieux était protégé par la législation chinoise en matière de droit d'auteur, le tribunal a également examiné l'éligibilité à une protection du rapport généré automatiquement par le logiciel. Dans son analyse de la protection des œuvres exclusivement générées par une IA, le tribunal estime que la notion de paternité exige que l'œuvre soit créée par une personne physique. Par ailleurs, le tribunal procède à une analyse intéressante des mesures incitatives et rejette la conclusion selon laquelle l'œuvre devrait être librement disponible dans le domaine public. Par conséquent, le tribunal considère qu'une certaine protection devrait être accordée à l'utilisateur - et non au développeur de logiciel déjà rémunéré par des droits d'auteur sur le logiciel - afin d'encourager l'achat du logiciel ainsi que la production et la distribution des œuvres. Malheureusement, le jugement ne précise pas sous quelle forme cette protection doit être mise en place.

Dans une décision ultérieure concernant l'affaire *Shenzhen Tencent c. Yinxun*, le tribunal d'instance de Nanshan, à Shenzhen, a confirmé en substance le jugement de Beijing⁴⁷¹. Les deux décisions sont concordantes dans la mesure où les tribunaux ont protégé les contributions originales des agents humains, plutôt que la créativité exclusivement générée par l'IA. La demanderesse, Tencent Technology, a développé un assistant d'écriture doté d'une IA, *Dreamwriter*. En août 2018, la demanderesse a publié sur son site internet l'une des œuvres créées par cette IA en informant les lecteurs que l'article avait été écrit par l'IA *Dreamwriter* de Tencent. La défenderesse est présumée avoir publié cet article sur son site internet sans l'autorisation de la demanderesse. La demanderesse a entamé une action en justice pour contrefaçon en faisant valoir qu'en sa qualité d'auteur de l'article, elle détenait des droits exclusifs en vertu de la législation sur le droit d'auteur. Elle affirme que l'article a été généré sous son contrôle et qu'elle est responsable de son organisation et de sa création, ainsi que de toute responsabilité en découlant. Au bénéfice de la demanderesse, le tribunal a établi que l'article remplissait les conditions pour être une œuvre littéraire originale, puisque le contenu était le produit des données d'entrée, des conditions de déclenchement et de la disposition des modèles et des ressources sélectionnés par une entité opérationnelle de la demanderesse. Étant donné que l'expression de l'article provient des choix et des dispositions individuels faits par la demanderesse, l'article généré par IA a été considéré comme une œuvre de personnes morales en vertu de l'article 11 de la loi sur le droit d'auteur, et la défenderesse a été tenue pour responsable d'une violation du droit d'auteur. Cependant, même si le tribunal a pu considérer l'œuvre comme une création intellectuelle intégrée, provenant à la fois de la contribution d'une équipe humaine et de l'exploitation de *Dreamwriter*, la protection accordée découle apparemment davantage de la contribution de l'équipe humaine que d'un quelconque apport de l'IA.

⁴⁷¹ *Shenzhen Tencent c. Yinxun*, tribunal d'instance populaire de Nanshan, Shenzhen, Province de Guangdong [2019] n° 14010 (Chine), disponible sur : <https://mp.weixin.qq.com/s/jjv7aYT5wDBldTVWXV6rdQ>. He K. (2020a), « Another decision on AI-generated work in China : Is it a Work of Legal Entities », *The IPKat*, <http://ipkitten.blogspot.com/2020/01/another-decision-on-ai-generated-work.html>.

5.3.3. Originalité : une machine peut-elle être originale ?

Parallèlement à la conception de la notion de paternité, la notion d'originalité en tant que condition à la protection au titre du droit d'auteur, semble également exclure la protection de la créativité artificielle. La doctrine et la jurisprudence interprètent l'originalité sur la base d'un modèle anthropocentrique qui met l'accent sur la conscience de soi. L'originalité est définie par une approche personnaliste qui considère qu'une œuvre est originale si elle représente la personnalité de l'auteur. Le terme même d'« auteur » véhicule cette acception, puisque l'étymologie la plus répandue le rattache au terme *αὐτός* en grec ancien, qui signifie « soi »⁴⁷². Cette caractérisation de l'originalité s'appuie sur les théories personnalistes, selon lesquelles les productions intellectuelles sont des manifestations ou des extensions de la personnalité de leurs créateurs⁴⁷³. Par conséquent, l'originalité en tant que représentation de « soi » et de la conscience de soi est théoriquement hors de la portée des créations générées automatiquement. Cette conception de l'originalité a été largement adoptée par la majorité des juridictions. Elle a écarté les approches précédentes basées sur la théorie de l'équité de Locke et sur les doctrines valorisant « la sueur du front » ainsi que « le savoir-faire, le travail et l'effort » dans la création d'une œuvre intellectuelle, indépendamment du fait que l'œuvre soit représentative ou non de la personnalité de l'auteur⁴⁷⁴.

Au sein de l'Union européenne, trois directives ont harmonisé verticalement la notion d'originalité. Selon l'article 1, paragraphe 3 de la Directive sur la protection des programmes d'ordinateur, l'article 6 de la Directive sur la durée de protection du droit d'auteur et l'article 3, paragraphe 1 de la Directive sur la protection des bases de données, une œuvre est réputée originale si elle est « une création intellectuelle propre à son auteur »⁴⁷⁵. Par la suite, la CJUE a étendu « horizontalement » l'originalité à toutes les questions relatives au droit d'auteur et clarifié quelque peu la portée de la notion. Dans l'affaire *Infopaq*, la CJUE note que « [c]e n'est qu'à travers le choix, la disposition et la combinaison de ces mots qu'il est permis à l'auteur d'exprimer son esprit créateur de manière originale et d'aboutir à un résultat constituant une création intellectuelle⁴⁷⁶. » Dans l'affaire *Eva-Maria Painer*, l'arrêt spécifie en outre qu'une œuvre - en l'espèce, une photographie de portrait - est originale et peut être protégée, si :

- 1) elle est une création intellectuelle de l'auteur ;

⁴⁷² Frosio G., *Reconciling Copyright with Cumulative Creativity: The Third Paradigm*, Edward Elgar, Cheltenham, p. 16.

⁴⁷³ Fichte J., « Proof of the illegality of reprinting: a rationale and a parable », *Berlinische Monatsschrift* 21, p. 447. Hegel G.H., *Philosophy of rights*, Thomas Knox, Clarendon Press, Oxford, paragraphe 69. Kant I. (1785), « Von der Unrechtmäßigkeit des Büchernachdrucks » [*De l'illégitimité de la contrefaçon des livres*], *Berlinische Monatsschrift* 5, pp. 403-417. Voir également Fisher W., « Theories of intellectual property », in Munzer S. (éd.) *New essays in the legal and political theory of property*, Cambridge University Press, Cambridge, pp. 168-200.

⁴⁷⁴ Voir par exemple *International News Service c. Associated Press*, 248 U.S. 215 (1918) (États-Unis). *Jeweler's Circular Publishing Co. c. Keystone Publishing Co.*, 281 F. 83 (2^e circ. 1922) (États-Unis). Voir également Rahmatian A., « Originality in UK Copyright Law The Old "Skill and Labour" Doctrine Under Pressure », *IIC* 44, pp. 4-34.

⁴⁷⁵ Voir sur ce point l'argumentation de Rosati E., *Originality in EU Copyright - Full Harmonization through Case Law*, Edward Elgar, Cheltenham.

⁴⁷⁶ *Infopaq International A/S c. Danske Dagblades Forening*, C-5/08 (2009) ECLI:EU:C:2009:465.

- 2) elle reflète la personnalité de ce dernier ;
- 3) elle se manifeste par les choix libres et créatifs de celui-ci lors de la réalisation de cette photographie⁴⁷⁷.

À travers ces différents choix, l'auteur d'une photographie de portrait est ainsi en mesure d'imprimer sa « touche personnelle » à l'œuvre créée⁴⁷⁸. Enfin, dans l'affaire *Football Dataco*, la CJUE rejette toute doctrine se référant à la « sueur du front » et établit que le travail et le savoir-faire significatifs requis de la part de l'auteur ne sauraient, en tant que tels, justifier une protection au titre du droit d'auteur s'ils n'expriment aucune originalité dans le choix ou la disposition⁴⁷⁹. Les œuvres produites uniquement sur la base de règles ou de contraintes techniques n'ont pas la liberté créatrice requise pour bénéficier d'une paternité.

La jurisprudence américaine a également entériné cette approche de l'originalité fondée sur la personnalité. Depuis les premières affaires telles que l'affaire *Burrow-Giles c. Sarony* qui portait sur la possibilité de protéger les droits d'auteur d'une photographie de portrait d'Oscar Wilde, la Cour suprême des États-Unis a établi que l'originalité découlait des choix créatifs libres de l'auteur qui imprègnent l'œuvre de sa personnalité⁴⁸⁰ « de telle sorte que le produit final reproduit sa conception et sa vision » de ce que l'œuvre devrait être⁴⁸¹. En particulier, dans l'affaire *Burrow-Giles*, le tribunal a jugé que les photographies étaient protégées par le droit d'auteur parce qu'elles portaient l'empreinte de la « propre conception mentale originale » du photographe⁴⁸². Plus tard, dans l'affaire *Feist c. Rural*, la Cour suprême des États-Unis a clairement établi que seules les œuvres dotées d'un minimum de créativité représentant la personnalité de l'auteur pouvaient être originales et que le travail et les efforts fournis à eux seuls pour créer une œuvre ne sauraient prétendre à une protection au titre du droit d'auteur⁴⁸³. À partir de ces considérations systémiques, il a été établi que les produits tels qu'un code informatique⁴⁸⁴ ou un inventaire de pièces matérielles numérotées automatiquement et créées à l'aide de systèmes logiciels n'avaient pas l'originalité requise pour être protégés par le droit d'auteur⁴⁸⁵.

En fait, P. Samuelson et plusieurs autres auteurs affirment qu'il n'y a pas de barrières légales aux États-Unis empêchant de traiter une machine comme un auteur, car « [l]e niveau d'originalité requis pour le droit d'auteur est suffisamment bas pour que les œuvres générées par ordinateur, même si elles sont créées exclusivement par une machine, puissent bénéficier de cette protection »⁴⁸⁶. En réponse, je dirais qu'au vu de l'affaire *Feist*, la question de l'originalité ne se pose pas seulement en termes quantitatifs. Aux fins de

⁴⁷⁷ *Eva-Maria Painer c. Standard VerlagsGmbH et autres*, C-145/10 (2011) ECLI:EU:C:2011:798.

⁴⁷⁸ *Eva-Maria Painer c. Standard VerlagsGmbH et autres*, C-145/10 (2011) ECLI:EU:C:2011:798.

⁴⁷⁹ *Football Dataco Ltd et autres c. Yahoo! UK Ltd et autres*, C-604/10 (2012) ECLI:EU:C:2012:115.

⁴⁸⁰ *Burrow-Giles Lithographic Co. c. Sarony*, 111 U.S. 53 (1884) (États-Unis).

⁴⁸¹ *Lindsay c. The Wrecked and Abandoned Vessel R.M.S. Titanic*, 52 U.S.P.Q.2d 1609 (S.D.N.Y. 1999) (États-Unis).

⁴⁸² *Burrow-Giles Lithographic Co. c. Sarony*, 111 U.S. 53 (1884) (États-Unis).

⁴⁸³ *Feist Publications c. Rural Telephone Service*, 499 U.S. 340 (1991) (États-Unis).

⁴⁸⁴ *Brief English Systems c. Owen*, 48 F.2d 555 (2^e circ. 1931) (États-Unis).

⁴⁸⁵ *Southco, Inc. c. Kanebridge Corporation*, 390 F.3d 276 (3^e circ. 2004) (États-Unis).

⁴⁸⁶ Samuelson P. (1986), *op.cit.*, pp. 1199-1200. Voir également Brown N., « Artificial Authors: A case for copyright in computer-generated works », *Columbia Science and Technology Law Review* 9, pp. 24-27. Kaminski M., « Authorship, Disrupted: AI Authors in Copyright and First Amendment Law », *UC Davis Law Review* 51, p. 601. Pour un avis contraire, voir Clifford, R.D. (1996), *op.cit.*, pp. 1694-1695.

l'appréciation de la créativité artificielle, peu importe que le niveau d'originalité soit faible ou élevé. La norme que l'IA n'est pas en mesure d'atteindre est d'ordre qualitatif et non pas quantitatif. L'IA ne peut pas exprimer de « soi ». La créativité qu'elle génère ne peut pas exprimer la personnalité de l'auteur puisque l'IA est dénuée de personnalité. À cet égard, l'adhésion des États-Unis à la Convention de Berne en 1988 et l'affaire *Feist* en 1991 marquent la cristallisation d'une vision globale, plus harmonisée, du droit d'auteur. Cet alignement des États-Unis sur le modèle européen implique également une conception de l'originalité fondée sur la théorie personnaliste⁴⁸⁷.

Plus récemment, quelques juridictions encore divergentes - majoritairement dans les pays de *common law* - ont rejoint cette approche de l'originalité fondée sur la personnalité. C'est le cas de l'Australie⁴⁸⁸, de l'Inde⁴⁸⁹ et du Royaume-Uni⁴⁹⁰, qui ont finalement rejeté les approches précédentes fondées sur le « travail, le savoir-faire et l'effort ». Seuls quelques pays, dont l'Afrique du Sud⁴⁹¹ et la Nouvelle-Zélande⁴⁹², suivent encore les doctrines de la « sueur du front » et rejettent les approches de l'originalité fondées sur la personnalité.

Pour résumer, il semble y avoir une interprétation internationale extrêmement cohérente de la notion d'originalité qui met l'accent sur une vision anthropocentrique selon laquelle une œuvre est originale si elle est une représentation de « soi », c'est-à-dire une représentation de la personnalité de l'auteur. Ce n'est que si ce lien intime entre l'auteur et l'œuvre est présent que la condition d'originalité est remplie et que la protection est accordée. Bien entendu, seul un être sensible et conscient de soi-même est capable de produire une représentation de « soi » à travers une œuvre. En revanche, même si l'on fait abstraction de toute interprétation doctrinale anthropocentrique de la paternité, en l'absence de conscience de soi du créateur, l'exigence d'originalité qui réside dans la représentation de la personnalité de l'auteur n'est jamais satisfaite. Par conséquent, à moins que l'on puisse affirmer que les machines ont développé une conscience de soi, ce qui pourrait être le cas avec une IA futuriste potentiellement performante mais certainement pas de nos jours⁴⁹³, la créativité artificielle n'est pas en mesure de répondre à l'exigence d'originalité inscrite dans le cadre juridique actuel⁴⁹⁴.

Conformément aux dires de certains, seule une nouvelle approche, peut-être plus formelle et plus objective - par opposition à l'approche subjective actuelle - du concept d'originalité pourrait inclure dans le champ d'application de la protection du droit d'auteur les œuvres créées par des robots créatifs ainsi que les œuvres d'art générées par des outils

⁴⁸⁷ Price M.E. & Pollack M., *The Author in Copyright: Notes for the Literary Critic*, 10 *Cardozo Arts & Entertainment Law Journal* 703 (1992), pp. 717-720, <https://larc.cardozo.yu.edu/faculty-articles/123>.

⁴⁸⁸ *IceTV Pty Ltd c. Nine Network Australia Pty Ltd* [2009] HCA 14, 239 CLR 458 (AUS).

⁴⁸⁹ *Eastern Book Co. & Ors c. D.B. Modak & Anr* (2008) 1 SCC 1 (Inde).

⁴⁹⁰ *Temple Island Collections c. New English Teas* (n° 2) [2012] EWPC 1. Rahmatian A., *op.cit.* pp. 4-34

⁴⁹¹ *Appleton c. Harnischfeger Corp.* 1995 (2) SA 247 (AD), pp. 43-44 (Afrique du Sud).

⁴⁹² *Henkel KgaA c. Holdfast* [2006] NZSC 102, [2007] 1 NZLR 577 (NZ).

⁴⁹³ Zimmerman E., *op.cit.* pp. 14-21.

⁴⁹⁴ Clifford, R.D. (1996), *op.cit.*, 1694-1695. Deltorn J.M. (2017), *op.cit.*, p. 7. Deltorn J.M. et Macrez F. (2018), *op.cit.*, p. 8. Gervais D.J., « The Machine as Author », *Iowa Law Review*, vol. 105, pp. 1-60. Mezei P., *From Leonardo to the Next Rembrandt - The Need for AI-Pessimism in the Age of Algorithms*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3592187. Ramalho A., *op.cit.* pp. 22-24.

numériques⁴⁹⁵. En partant de ce point de vue objectif, un juge devrait alors examiner le résultat final *per se*, en se référant au domaine de l'art, à l'opinion objective des utilisateurs et à la similarité avec d'autres œuvres, tout en ignorant l'intention subjective de l'auteur⁴⁹⁶. À cet égard, la norme d'originalité au regard du droit d'auteur devrait s'aligner davantage sur la norme de nouveauté du droit des brevets, qui considère les objets protégeables dans une optique sociale/historique plutôt que dans une perspective individuelle/subjektive⁴⁹⁷.

5.4. Options politiques : des incitations sont-elles nécessaires ?

Certains experts et tribunaux font valoir que du point de vue de la « théorie incitative », un système juridique qui n'accorde pas de protection à la créativité artificielle produit des effets négatifs. La théorie incitative, ou théorie utilitariste⁴⁹⁸, qui prédomine aux États-Unis et dans les juridictions de *common law*, est plus éloignée de la dimension humaine de l'auteur que les théories personnalistes qui influencent fortement les juridictions de droit civil⁴⁹⁹. Cela laisse plus de place aux arguments en faveur d'une paternité non humaine et d'une protection de la créativité artificielle. Selon l'approche fondée sur la théorie incitative, « fournir des incitations financières afin d'encourager la croissance et le développement de l'industrie de l'IA et d'assurer la diffusion des œuvres générées par IA pourrait être la finalité suprême de l'octroi des droits d'auteur aux auteurs humains⁵⁰⁰. » Sachant qu'un ordinateur n'a pas besoin d'une incitation pour produire un résultat, l'incitation peut être utile à la personne qui collabore avec l'ordinateur⁵⁰¹. En particulier, certains auteurs soutiennent qu'il devrait y avoir une incitation supplémentaire pour encourager l'industrie à investir le temps et l'argent nécessaires pour apprendre aux machines à se comporter intelligemment⁵⁰² ou pour récompenser les utilisateurs qui entraînent et instruisent l'IA pour qu'elle génère des contenus⁵⁰³. D'aucuns affirment que, du point de vue utilitariste, les considérations d'ordre politique rendent incontournable

⁴⁹⁵ Yanisky-Ravid S. et Velez-Hernandez L.A., « Copyrightability of Artworks Produced by Creative Robots, Driven by Artificial Intelligence Systems and the Originality Requirement The Formality-Objective Model », *Minnesota Journal of Law, Science & Technology* 19(1), pp. 40-48.

⁴⁹⁶ Bonadio E. et McDonagh L., « Artificial Intelligence as Producer and Consumer of Copyright Works: Evaluating the Consequences of Algorithmic Creativity », *Intellectual Property Quarterly*, 2, pp. 112-137, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3617197.

⁴⁹⁷ Cf. Boden M., *The Creative Mind: Myths And Mechanisms*, Routledge, Londres, p. 32.

⁴⁹⁸ Fisher W., *op.cit.* pp. 177-180.

⁴⁹⁹ Kaminski M., *op.cit.*, p. 599.

⁵⁰⁰ Hristov K., « Artificial Intelligence and the Copyright Dilemma », *IDEA: The Intellectual Property Law Review* 57, p. 444. Voir également Brown N., *op.cit.*, 20-21.

⁵⁰¹ Hristov K., *op.cit.*, 438-439. Miller A.R., *op.cit.*, 1067.

⁵⁰² Bridy A. (2012), *op.cit.*, 1-27. Butler T., *op.cit.*, p. 735. Farr E.H., « Copyrightability of Computer-Created Works », *Rutgers Computer & Technology Law Journal* 15, pp. 73-74. Kasap A., *op.cit.* pp. 361-364 ; Abbott R., « I Think, Therefore I Invent: Creative Computers and the Future of Patent Law », 57 B.C.L. Rev. 1079 (2016), <http://lawdigitalcommons.bc.edu/bclr/vol57/iss4/2>. Milde K.F., *op.cit.*, p. 390.

⁵⁰³ Brown N., *op.cit.*, p.37. Denicola R., *op.cit.*, p. 283. Ralston W.T., « Copyright in Computer-Composed Music: HAL Meets Handel », *Journal of the Copyright Society of the USA* 52, pp. 303-304. Samuelson P. (1986), *op.cit.* pp. 1224-1228.

l'octroi d'une certaine forme de protection des produits générés par IA, qu'il s'agisse d'une protection au titre du droit d'auteur ou du droit de la concurrence, ou d'une protection *sui generis*⁵⁰⁴, puisque le processus de création du produit – que ce soit par l'humain ou par l'ordinateur – n'a pas d'incidence sur sa contribution au bien-être public⁵⁰⁵.

Néanmoins, la plupart des juridictions de droit civil ne sont pas tant imprégnées des considérations liées au bien-être et à l'incitation et peuvent privilégier l'équilibre systémique, préférant suivre l'approche théorique fondée sur la personnalité qui détermine la perspective du droit d'auteur dans le droit civil et, *de facto*, la notion d'originalité en vigueur dans la grande majorité des juridictions. Par conséquent, bien que les créations générées par IA puissent justifier des incitations pour soutenir l'innovation et la commercialisation, la nécessité de telles incitations est discutable si l'on considère l'impact qu'elles peuvent avoir sur les créations humaines⁵⁰⁶. Par exemple, étant donné le grand nombre de créations automatisées, le fait d'accorder une protection à ces œuvres pourrait dévaloriser la paternité humaine et les emplois existants dans ce domaine et entraver la créativité⁵⁰⁷, car cela risque de décourager les artistes de publier leurs créations par crainte de porter atteinte aux œuvres protégées⁵⁰⁸ ou d'encombrer l'écosystème créatif avec des productions standardisées et uniformisées générées par l'IA, ce qui aurait un impact sur la politique en matière d'identité et de diversité culturelle.

Il reste à déterminer s'il est utile d'étendre la protection actuelle du droit d'auteur aux œuvres générées par ordinateur. Le cadre juridique en vigueur offre d'ores et déjà une protection suffisante par le biais du droit des brevets et du droit d'auteur pour les logiciels sous-jacents, de la protection *sui generis* des bases de données ou d'autres dispositifs juridiques, tels que le droit de la concurrence, pour protéger les œuvres automatisées, sans qu'il soit nécessaire d'étendre le régime actuel du droit d'auteur aux auteurs non humains⁵⁰⁹. Il convient d'aborder les différents aspects du problème dans une perspective juridique et économique avant de se prononcer en faveur de l'une ou l'autre des solutions⁵¹⁰. Ginsburg et Budiardjo insistent bien sur ce point : « [n]ous pouvons évoquer toute une série de scénarios pour soutenir ou contester l'appel à une protection *sui generis*, mais en l'absence de preuves empiriques, il serait imprudent (et prématuré) de chercher à concevoir un régime couvrant les productions sans auteur⁵¹¹. »

⁵⁰⁴ Milde K.F., *op.cit.* pp. 400-403.

⁵⁰⁵ Butler T., *op.cit.*, p. 735. Denicola R., *op.cit.*, p. 273. Kaminski M., *op.cit.*, p. 599.

⁵⁰⁶ Craglia M., *op.cit.* pp. 67-68.

⁵⁰⁷ Bonadio E. et McDonagh L., *op.cit.*

⁵⁰⁸ Deltorn J.M. (2017), *op.cit.*

⁵⁰⁹ Deltorn J.M. et Macrez F. (2018), *op.cit.*, p. 24.

⁵¹⁰ Craglia M., *op.cit.*, p. 68.

⁵¹¹ Ginsburg J. et Budiardjo L.A., *op.cit.*, p. 448. Voir également Ginsburg J., *op.cit.* pp. 131-135.

5.4.1. Absence de protection : les œuvres créées par IA dans le domaine public

Il ressort de notre analyse précédente des conditions d'éligibilité à la protection que la conception des notions de personnalité juridique, de paternité et d'originalité dans le cadre du régime actuel du droit d'auteur semble exclure la créativité artificielle du champ de protection au titre du droit d'auteur⁵¹². La relégation de la créativité artificielle dans le domaine public serait donc un choix politique non seulement possible, mais également le plus indiqué dans le cadre juridique actuel.

Selon cette approche, la titularité du droit d'auteur dépend de la teneur de l'intervention humaine. La simple sélection et classification des données par des humains ne suffit pas à satisfaire au critère d'« originalité » ; en revanche, une contribution humaine effective et substantielle pour guider le système d'IA dans sa création est nécessaire pour l'octroi de la protection⁵¹³. Ce n'est que lorsqu'il y a un apport humain substantiel et que tous les choix créatifs sont intégrés dans le code informatique ou le manuel d'utilisation que le droit d'auteur peut être dévolu à l'auteur humain⁵¹⁴. À cet égard, quatre modèles d'attribution de la paternité ont été identifiés : (1) la paternité exclusive de l'utilisateur de l'outil, si le concepteur de l'outil ne contribue pas au travail créatif généré ; (2) la paternité exclusive des concepteurs de l'outil, si l'opérateur ne joue aucun rôle dans le produit et si l'outil auto-générateur crée un produit basé sur l'apprentissage et la matière première créative fournis par le concepteur ; (3) la paternité conjointe de l'utilisateur et du programmeur, lorsque le produit reflète les contributions créatives du concepteur et de l'utilisateur ; (4) les œuvres sans auteur : ni le concepteur ni l'utilisateur ne contribuent suffisamment à l'expression du résultat pour produire une œuvre originale⁵¹⁵.

En tout état de cause, si la production créative résulte à la fois de choix humains et mécaniques, les éléments résultant de choix mécaniques doivent être exclus, comme il est d'usage pour les œuvres du domaine public⁵¹⁶. Seules les contributions humaines indépendantes et relevant du droit d'auteur pourront être protégées.

Un test supplémentaire a été proposé pour déterminer si une œuvre méritait d'être protégée en fonction de l'importance de l'empreinte transmise à l'œuvre finale par les développeurs et de la capacité de l'utilisateur ou du développeur à prévoir le résultat⁵¹⁷. Un tel test impliquerait la mise en œuvre des exigences traditionnelles du droit d'auteur n'accordant une protection que si l'œuvre est le produit de l'imagination d'auteurs humains

⁵¹² Aplin T. et Pasqualetto G., *op.cit.*, paragraphe 5.01-09. Clifford, R.D. (2018), « Creativity Revisited », *IDEA The IP Law Review*, 59, pp. 26-29. Clifford, R.D. (1996), *op.cit.*, 1700-1702. Huson G., « I, Copyright », *Santa Clara High Technology Law Journal* 35, pp. 72-78. Mezei P., *op.cit.* Palace V.M., « What if Artificial Intelligence Wrote This: Artificial Intelligence and Copyright Law », *Florida Law Review* 71(1), pp. 238-241. Saiz Garcia C., « Las obras creadas por sistemas de inteligencia artificial y su protección por el derecho de autor » *InDret* 1, pp. 38-39, <https://ssrn.com/abstract=3365458>. Gervais D.J., *op.cit.* pp. 1-60. Ramalho A., *op.cit.* pp. 22-24. Svedman M., *op.cit.* pp. 1-22 ;

⁵¹³ Selvadurai N. et Matulionyte R., « Reconsidering creativity: copyright protection for works generated using artificial intelligence », *Journal of Intellectual Property Law & Practice* jpaa062, p. 539.

⁵¹⁴ Gervais D.J., *op.cit.* pp. 51-60. Selvadurai N. et Matulionyte R., *op.cit.*, p. 538.

⁵¹⁵ Ginsburg J. et Budiardjo L.A., « Authors and Machines », *Berkeley Technology Law Journal* 34(2), pp. 404-445.

⁵¹⁶ Gervais D.J., *op.cit.*, p. 54.

⁵¹⁷ Boyden B., « Emergent Works », *Columbia Journal of Law and the Arts* 39, pp. 377-394.

et de leur conception de celle-ci. Si ce test est mis en œuvre, la législation sur le droit d'auteur devrait s'orienter vers un déplacement de la charge de la preuve traditionnelle, en ce sens que le demandeur devra prouver la paternité humaine de certains produits générés par IA en établissant que le produit est conçu en amont pour intégrer un sens ou un message que l'auteur souhaite transmettre à son public⁵¹⁸.

5.4.2. Paternité et fictions juridiques : l'auteur doit-il être un humain ?

Afin d'éviter que la créativité artificielle ne tombe dans le domaine public et d'accorder les incitations nécessaires aux agents humains impliqués dans le processus de création de l'IA, des propositions ont été faites et une loi a été adoptée pour mettre en place une fiction juridique, de sorte que la paternité des œuvres créées par IA soit conférée aux agents qui fournissent le savoir-faire, le travail et l'effort permettant en premier lieu de créer, d'entraîner ou d'instruire l'IA. Cette approche est appelée la « théorie de l'auteur humain fictif »⁵¹⁹.

Cette démarche politique est apparue assez tôt, alors que le potentiel créatif et les mécanismes de l'apprentissage machine et de l'IA étaient totalement inconnus. En 1985, au Royaume-Uni, la protection par le droit d'auteur d'une séquence créée par ordinateur pour une loterie faisait déjà débat lors de l'affaire *Express Newspapers c. Liverpool Daily Post*. Justice Whitford a attribué à la demanderesse la protection de la séquence automatisée au titre du droit d'auteur et rejeté l'idée que le droit d'auteur attaché à l'œuvre puisse être dévolu à l'ordinateur. Selon le juge, l'ordinateur est un simple outil de création et le fait d'affirmer que l'ordinateur est l'auteur revient à dire que dans une œuvre écrite, « c'est le stylo qui est l'auteur de l'œuvre et non pas la personne qui tient le stylo »⁵²⁰. Dernièrement, cette opinion a été synthétisée avec pertinence par Dan Burk en termes de causalité proche, d'intention et de volonté, en ce sens que « [s]'il y a un auteur, c'est l'un ou plusieurs des humains qui sont suffisamment proches, en termes de causalité, de la production du résultat. [...] Mais l'auteur n'est jamais la machine. »⁵²¹.

Le Royaume-Uni a été la première juridiction à aménager une protection spécifique pour la créativité générée par ordinateur⁵²². L'article 9, paragraphe 3 du *Copyright Designs and Patents Act 1988* (loi de 1988 sur les droits d'auteur, les dessins et modèles et les brevets

⁵¹⁸ Boyden B., *op.cit.*, 393-394.

⁵¹⁹ Wu A.J., « From Video Games to Artificial Intelligence: Assigning Copyright Ownership to Works Generated by Increasingly Sophisticated Computer Programs », *AIPLA Quarterly Journal*, pp. 173-174.

⁵²⁰ *Express Newspapers Plc c. Liverpool Daily Post & Echo Plc* [1985] 1 WLR 1089 (Royaume-Uni).

⁵²¹ Burk D.L., « Thirty-Six Views of Copyright Authorship, By Jackson Pollock », *Houston Law Review* 58, pp. 1-38. Voir aussi Hedrick S.F., « I 'Think', Therefore I Create: Claiming Copyright in the Outputs of Algorithms », *NYU Journal of Intellectual Property & Entertainment Law* 8(2), pp. 324-375, et Grimmelmann J., « There is No Such Thing as a Computer-Authored Work And It is a Good Thing, Too », *Columbia Journal of Law and the Arts* 39, pp. 403-416.

⁵²² Guadamuz A., « Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works », *Intellectual Property Quarterly*, pp. 169-186.



- CDPA)⁵²³ spécifie que pour les œuvres générées par ordinateur, l'auteur est la personne qui prend les dispositions nécessaires à la création de l'œuvre. En outre, l'article 178 dispose que « la production par ordinateur, en relation avec une œuvre, signifie que l'œuvre est produite par ordinateur dans des circonstances telles qu'il n'y a pas d'auteur humain de l'œuvre. » Dans le cadre de ce régime, la durée de protection des œuvres générées par ordinateur serait de 50 ans à compter du moment où l'œuvre a été produite. Peu après, d'autres pays de *common law*, dont Hong Kong, l'Inde, l'Irlande, Singapour et la Nouvelle-Zélande, ont adopté des dispositions juridiques similaires⁵²⁴.

Il y a toutefois deux problèmes qui remettent en cause ce dispositif. Le premier est d'ordre fondamental et systémique. Cette approche serait-elle viable dans un cadre juridique qui s'appuie sur la notion d'originalité en tant qu'expression de la personnalité de l'auteur, telle qu'elle est adoptée par le droit de l'UE ainsi que par la plupart des juridictions internationales ? Il est clair que la programmation, l'entraînement et la transmission d'instructions ne seraient probablement pas en mesure de satisfaire à l'exigence d'une contribution originale émanant des humains, puisque toute « expression » serait en définitive le résultat du processus créatif de l'IA. Tant que la norme subjective actuelle en matière d'originalité sera en vigueur, toute théorie d'un auteur humain fictif s'effondrera devant le défaut d'originalité de la créativité artificielle. L'œuvre elle-même, dont la paternité fictive est attribuée à un agent humain, resterait en fait sans originalité, donc sans protection. Il convient de noter que la « théorie de l'auteur humain fictif » et les approches basées sur la prise en charge des « dispositions nécessaires » ont été adoptées au Royaume-Uni et dans d'autres pays de *common law* alors que les normes d'originalité fondées sur la « la sueur du front » ou « le savoir-faire et le travail » étaient encore largement en vigueur dans ces juridictions. Depuis lors, comme indiqué précédemment, des changements sont intervenus, le critère de la personnalité pour établir l'originalité remplaçant totalement ou partiellement toute autre approche⁵²⁵ et remettant en cause l'approche systémique de l'article 9, paragraphe 3 de la CDPA.

La deuxième question est d'ordre plus concret et pointe le fait que, dans une telle approche, il est compliqué d'établir quelle est la personne responsable des dispositions nécessaires⁵²⁶. L'œuvre créée par IA appartient-elle à la personne qui a produit le système, comme le développeur du logiciel, le fabricant, la personne qui l'a entraîné, ou à la personne qui lui a fourni des entrées spécifiques, telle que l'utilisateur⁵²⁷ ? Selon M. Guadamuz, l'ambiguïté du système devrait en fait être considérée comme un élément positif

⁵²³ Copyright, Designs and Patents Act 1988 (loi de 1988 sur le droit d'auteur, les dessins et modèles et les brevets). Disponible sur : <https://www.legislation.gov.uk/ukpga/1988/48/contents> (Royaume-Uni)

⁵²⁴ Copyright Ordinance cap 528, section 11(3), <https://www.elegislation.gov.hk/hk/cap528> (Hong Kong) ; Copyright Act 1957, section 2(d)(vi), <https://www.wipo.int/edocs/lexdocs/laws/en/in/in122en.pdf> (Inde) ; Copyright and Related Rights Act 2000, section 21(f), <http://www.irishstatutebook.ie/eli/2000/act/28/enacted/en/html> (Irlande) ; Copyright Act 1987, chapitre 63, section 7A, <https://sso.agc.gov.sg/Act/CA1987#pr27-> (Singapour) ; Copyright Act 1994, section 5(2), <http://www.legislation.govt.nz/act/public/1994/0143/latest/DLM345634.html> (Nouvelle-Zélande).

⁵²⁵ *Temple Island Collections c. New English Teas* (n° 2) [2012] EWPC 1. Guadamuz A., *op.cit.* pp. 178-180 ; Rahmatian A., *op.cit.* pp. 4-34.

⁵²⁶ Dorotheu E., « Reap the benefits and avoid the legal uncertainty : who owns the creations of artificial intelligence ? », *Computer and Telecommunications Law Review* 21, p. 85.

⁵²⁷ Bonadio E. et McDonagh L., *op.cit.* pp. 117-119. Kasap A., *op.cit.* pp. 364-376.

qui estompe la dichotomie utilisateur/programmeur et permet une analyse au cas par cas⁵²⁸. En tout état de cause, la question visant à savoir « qui est responsable des dispositions nécessaires » a reçu de multiples réponses, tant par la jurisprudence que par la doctrine.

5.4.2.1. Le programmeur est-il l'auteur ?

Une première réponse possible à cette question nous a été fournie par l'affaire *Nova Productions c. Mazooma Games*, dans laquelle l'article 9, paragraphe 3 de la CDPA a été appliqué. L'affaire concernait les droits d'auteur liés à des images générées par un programme informatique utilisant des fichiers bitmap, qui s'affichaient à l'écran lorsque les utilisateurs jouaient à un jeu vidéo de billard. Le tribunal a refusé d'accorder la paternité des images à l'utilisateur au motif que sa contribution n'avait aucune dimension artistique :

L'apparence d'un écran particulier dépend dans une certaine mesure de la manière dont le jeu est joué. Par exemple, lorsqu'on tourne le bouton, la queue de billard tourne autour de la boule. De même, la puissance du coup est déterminée par le moment précis où le joueur choisit d'appuyer sur le bouton « play ». Cependant, le joueur n'est l'auteur d'aucun des dessins artistiques qui composent les cadres successifs (frames). Son apport n'est pas de nature artistique et il n'a fourni aucun savoir-faire, ni aucun travail de nature artistique. Il n'a pas, non plus, pris les dispositions nécessaires pour la création des cadres visuels. Son intervention se limite à jouer au jeu⁵²⁹.

En revanche, le tribunal identifie le programmeur comme le seul auteur et comme la personne ayant pris les dispositions nécessaires en précisant que « [l]es dispositions nécessaires à la création de l'œuvre ont été prises par [la demanderesse] puisqu'elle a conçu l'apparence des différents éléments du jeu, ainsi que les règles et la logique selon lesquelles chaque cadre est réalisé, et qu'elle a écrit le programme informatique correspondant.⁵³⁰ » En réalité, la séquence produite par Nova Productions pourrait découler directement de la technologie de base en cause, et, comme l'a fait valoir M. Guadamuz, une dévolution différente de la paternité pourrait en résulter selon les spécificités propres à l'affaire et à la technologie examinée⁵³¹. Néanmoins, cette approche a été proposée dans d'autres juridictions, notamment aux États-Unis, imposant aux tribunaux d'adapter les termes de la Copyright Act de telle sorte que, lorsque ni le programmeur, ni l'utilisateur ne satisfont aux exigences de la paternité d'une œuvre protégée par le droit d'auteur, le tribunal doive attribuer le droit d'auteur à celui qui détient le droit d'auteur sur le programme d'ordinateur⁵³².

L'attribution de la paternité au programmeur de contenus générant l'IA appelle une critique de fond. L'attribution de la paternité au développeur du logiciel - ou aux

⁵²⁸ Guadamuz A., *op.cit.* p. 177.

⁵²⁹ *Nova Productions Ltd c. Mazooma Games Ltd & Ors Rev 1* [2006] EWHC 24 (Ch) (20 janvier 2006) (Royaume-Uni). Voir également Farr E.H., *op.cit.*, 75-78.

⁵³⁰ *Nova Productions Ltd c. Mazooma Games Ltd & Ors Rev 1* [2006] EWHC 24 (Ch) (20 janvier 2006) (Royaume-Uni). Voir également Farr E.H., *op.cit.*, 73-74.

⁵³¹ Guadamuz A., *op.cit.*, p. 177.

⁵³² Wu A.J., *op.cit.* pp. 173-174.

propriétaires des technologies d'IA, tels que les entreprises et les investisseurs - constitue de toute évidence une erreur⁵³³. En fait, il ne semble pas, tout au moins dans la créativité basée sur les réseaux neuronaux de pointe (GAN et CAN), y avoir de lien de causalité direct entre les développeurs de logiciels et le produit final généré par l'IA, car l'expression inhérente au produit est le résultat de l'apprentissage de la machine et des instructions données pour créer ce produit particulier. Tout d'abord, à la lumière de ce qui précède et d'un point de vue systémique, on pourrait faire valoir que ce dispositif induit une incohérence fondamentale. En fait - comme l'a également souligné le tribunal de Beijing chargé des affaires en ligne dans une affaire mentionnée plus haut - le développeur de logiciels a déjà été rétribué par des droits exclusifs sur le logiciel qui génère les œuvres⁵³⁴. En outre, étant donné que cette fiction juridique est précisément censée être une incitation à créer des œuvres générées par l'IA, sachant que le statut d'œuvre relevant du domaine public n'aurait probablement pas cet effet incitatif, toute analyse économique sérieuse ne saurait que dissuader de faire un choix politique consistant à rétribuer deux fois le même acteur du marché. Enfin, d'un point de vue plus concret, cette orientation politique mettrait les développeurs en position de défense agressive de leurs droits d'auteur en faisant valoir ces derniers sur une multitude d'éléments de créativité,⁵³⁵ ce qui saperait toute motivation de la part des programmeurs d'origine pour créer davantage de logiciels⁵³⁶.

La jurisprudence américaine semble également confirmer la conclusion selon laquelle les logiciels et les produits sont deux entités très distinctes, et que la propriété des premiers ne saurait impliquer de droits sur les seconds. Dans deux affaires, les tribunaux ont décidé que le produit créé à l'aide de la copie illicite d'un logiciel ou d'un programme n'était pas considéré comme une œuvre dérivée illicite. Dans l'affaire *Design Data c. Unigate Enterprises*, la cour d'appel a confirmé la décision du tribunal d'instance et jugé que les droits d'auteur de la demanderesse sur le programme informatique ne s'étendaient pas aux produits créés par ledit programme (dessins et données pour des bâtiments en acier)⁵³⁷. Dans l'affaire *Rearden c. Walt Disney Co.*, les films créés au moyen d'une copie (illicite) du logiciel de la demanderesse ne sont pas considérés comme des œuvres dérivées du logiciel car, en dépit du fait que le logiciel a fourni une somme de travail substantielle, la majeure partie de l'expression créative présente dans le film est imputable à la défenderesse⁵³⁸.

⁵³³ Abbott R., « Artificial Intelligence, Big Data and Intellectual Property: protecting computer generated works in the United Kingdom » dans Aplin T. (éd.), *Research Handbook on intellectual property and digital technologies*, Edward Elgar, Angleterre, pp. 323-324. Svedman M., *op.cit.*, 10-11. Bridy A. (2016), « The Evolution of Authorship: Work made by Code », *Columbia Journal of Law and the Arts* 39, pp. 400-401. Bridy A. (2012), *op.cit.*, 24-25, paragraphe 62. Samuelson P. (1986), *op.cit.*, pp. 1207-1212.

⁵³⁴ *Beijing Feilin Law Firm c. Baidu Corporation* (26 avril 2019), Tribunal de Beijing chargé des affaires en ligne, (2018) Beijing 0491 Minchu n° 239. He K. (2020b), *op.cit.* Voir également Bonadio E. et McDonagh L., *op.cit.*, p. 117. Chen M., *op.cit.*, pp. 14-18. Samuelson P. (1986), *op.cit.*, pp. 1207-1212.

⁵³⁵ Svedman M., *op.cit.*, p. 14.

⁵³⁶ Huson G., *op.cit.*, p. 74.

⁵³⁷ *Design Data Corp. c. Unigate Enterprise*, n° 14-16701 (9^e circ. 2017) (États-Unis).

⁵³⁸ *Rearden c. Walt Disney Co.*, n° 17-cv-04006-JST (N.D. Cal. 2018).

5.4.2.2. L'utilisateur est-il l'auteur ?

L'attribution des droits sur un produit créé par IA à l'utilisateur du programme générateur pourrait être une solution plus judicieuse⁵³⁹. La récente décision de Beijing exposée plus haut s'appuie sur la même conclusion⁵⁴⁰. P. Samuelson fait valoir que l'utilisateur constitue la raison même pour laquelle une œuvre créée par IA voit le jour, et qu'il n'est donc pas « injuste, dans ces conditions, d'accorder certains droits à une personne qui utilise l'œuvre dans le but de créer des œuvres supplémentaires⁵⁴¹ ». Cette solution n'est pas une nouveauté au regard des normes en matière de droits d'auteur. Aux États-Unis, par exemple, les droits d'auteur - et la paternité - sont accordés aux utilisateurs au motif qu'ils sont l'instrument de fixation⁵⁴², au même titre qu'une personne qui enregistre une interprétation d'un morceau de jazz sur bande magnétique⁵⁴³. Dans un tel scénario, l'auteur de l'enregistrement sonore serait l'utilisateur et non pas les musiciens de jazz. De la même façon, l'utilisateur pourrait être considéré comme l'auteur de la fixation de l'œuvre générée par IA. Bien entendu, il conviendrait d'intégrer à cette fin une disposition spécifique, telle que l'article 9, paragraphe 3 de la CDPA. Il est probable que dans certains cas exceptionnels, comme, par exemple, lorsque l'utilisateur n'a pas d'autre contrôle sur le logiciel que de lancer son exécution, l'attribution du droit d'auteur à l'utilisateur serait un choix politique inopportun et en contradiction avec la théorie incitative du droit d'auteur⁵⁴⁴. En ce cas, une co-paternité entre les utilisateurs et les programmeurs pourrait éventuellement être une solution⁵⁴⁵, en fonction du régime juridique de co-paternité mis en place par les différentes juridictions.

Dans un « Projet de rapport sur les droits de propriété intellectuelle pour le développement des technologies de l'intelligence artificielle », le Parlement européen partage manifestement le même point de vue et propose de confier aux utilisateurs de l'IA les droits d'auteur sur les œuvres générées par l'IA. Le projet de rapport « considère que la protection des créations techniques et artistiques générées par l'IA doit être envisagée afin d'encourager cette forme de création ; considère que certaines œuvres générées par l'IA sont assimilables à des œuvres de l'esprit et que, dès lors, elles pourraient être protégées par le droit d'auteur⁵⁴⁶. » Par conséquent, le projet de rapport « recommande que la titularité des droits revienne à la personne qui édite et rend publique l'œuvre de manière licite à condition qu'un tel usage n'ait pas été expressément réservé par le concepteur de la technologie⁵⁴⁷. » Apparemment, cette proposition implique que la propriété des droits soit également cédée aux utilisateurs, puisque le projet de rapport fait référence à la protection

⁵³⁹ CONTU, *op.cit.*, p. 45. Voir également Ralston W.T., *op.cit.*, pp. 303-304.

⁵⁴⁰ *Beijing Feilin Law Firm c. Baidu Corporation* (26 avril 2019), Tribunal de Beijing chargé des affaires en ligne, (2018) Beijing 0491 Minchu n° 239. He K. (2020b), *op.cit.*

⁵⁴¹ Samuelson P. (2020), « AI Authorship? » *Communications of the ACM* 63(7), pp. 20-22. Samuelson P. (1986), « Allocating Ownership Rights in Computer-Generated Works », *University of Pittsburg Law Review* 47(4), pp. 1200-1204.

⁵⁴² 17 U.S.C. article 114.

⁵⁴³ Samuelson P. (1986), *op.cit.*, pp. 1200-1204. Cependant, la plupart des pays sont favorables à une protection des enregistrements sonores au titre des droits voisins, et non pas du droit d'auteur comme aux États-Unis.

⁵⁴⁴ Ralston W.T., *op.cit.*, pp. 304-305.

⁵⁴⁵ Par exemple, Bonadio E. et McDonagh L., *op.cit.*, 117-118.

⁵⁴⁶ Parlement européen (2020), *op.cit.*

⁵⁴⁷ Parlement européen (2020), *op.cit.*

du droit d'auteur et non pas à une protection *sui generis*, en spécifiant qu'« il est proposé d'évaluer l'opportunité d'attribuer les droits d'auteur d'une telle "œuvre créative" à la personne physique qui l'édite et la publie licitement [...] »⁵⁴⁸.

5.4.2.3. L'employeur est-il l'auteur ?

Certains experts ont proposé la doctrine du « *work-made-for-hire* » (travail sur commande - WMFH)⁵⁴⁹ comme cadre juridique pour garantir la protection, la paternité et la responsabilité des œuvres générées par les systèmes d'IA⁵⁵⁰. Ce modèle serait basé sur une fiction dans laquelle le système d'IA est un employé créatif ou un fournisseur indépendant des utilisateurs - personnes physiques ou morales - qui utilisent les systèmes d'IA et profitent de leurs avantages⁵⁵¹. Comme l'affirme P. Samuelson, « celui qui achète ou concède sous licence un programme générateur a en quelque sorte "employé" l'ordinateur et ses programmes pour ses activités créatives, selon des considérations similaires à celles qui sous-tendent la règle du travail sur commande et parlent en faveur de l'octroi aux utilisateurs des droits sur les œuvres créées par ordinateur⁵⁵². » Ainsi, la titularité du droit d'auteur et la responsabilité à l'égard de toute infraction découlant de l'œuvre incomberaient à la personne physique ou morale considérée comme l'employeur ou le commanditaire du système d'IA à l'origine de l'œuvre.

L'adoption d'une telle réglementation ne manquerait pas de susciter quelques critiques. Tout d'abord, on entend souvent dire que les employeurs sont traités comme les auteurs des œuvres réalisées sur commande, alors qu'ils ne jouent aucun rôle dans la production et que, par conséquent, des dispositions similaires pourraient être prises pour la créativité artificielle⁵⁵³. Il semble toutefois que cette position ne tienne pas compte du fait que, dans le cadre de la fiction juridique de la doctrine WMFH, l'œuvre sous-jacente a été créée par un auteur humain et répond au critère d'originalité dans le cadre juridique actuel, ce qui ne serait pas le cas avec la créativité artificielle. Par ailleurs, ce dispositif serait confronté à certains défis résultant du fait qu'il constitue une application abusive de la doctrine WMFH, sachant qu'il est pour le moins difficile de définir une relation juridique

⁵⁴⁸ Parlement européen (2020), *op.cit.*

⁵⁴⁹ En vertu de la législation américaine sur le droit d'auteur, le travail sur commande constitue une exception à la règle selon laquelle seul l'auteur (ou ceux qui en tirent des droits) peut revendiquer un droit d'auteur sur une œuvre protégeable. Selon cette doctrine, l'employeur est considéré comme l'auteur d'une œuvre même si c'est un employé qui a créé l'œuvre. Cette disposition est visée à l'article 101 du Copyright Act (loi sur le droit d'auteur). Une œuvre créée par un employé dans le cadre de son emploi ou commandée ou commissionnée spécialement pour être utilisée par l'employeur est une œuvre sur commande si les parties en conviennent ainsi dans un accord signé (United States Copyright Office, « *Works Made for Hire* ». Disponible sur : copyright.gov/circs/circ09.pdf).

⁵⁵⁰ Bridy A. (2016), *op.cit.*, pp. 400-401. Bridy A. (2012), *op.cit.*, 26, paragraphe 66. Hristov K., *op.cit.*, 431-454. Kaminski M., *op.cit.* Voir également Pearlman R., « Recognizing Artificial Intelligence (AI) as Authors and Inventors under U.S. Intellectual Property Law », *Richmond Journal of Law & Technology* 24, pp. 1-38. Yanisky-Ravid S. (2017), « Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era, The Human-Like Authors are Already Here: A New Model », *Michigan State Law Review*, pp. 659-726.

⁵⁵¹ Yanisky-Ravid S., *op.cit.*, pp. 659-726.

⁵⁵² Samuelson P. (2020), *op.cit.*, pp. 20-22 ; Samuelson P. (1986), *op.cit.*, pp. 1200-1204.

⁵⁵³ Brown N., *op.cit.*, p.39. Kaminski M., *op.cit.*, 602.

sous forme de contrat de travail ou de mandat entre un être humain et une machine⁵⁵⁴. L'application de la doctrine WMFH serait problématique en particulier dans des juridictions telles que celle de la France, où les transferts de paternité aux employeurs ou aux commanditaires doivent être explicitement prévus dans le contrat de travail ou de mandat et ne sauraient être implicites. Il semble évident que pour pouvoir appliquer la doctrine WMFH aux œuvres générées par l'IA, il faut impérativement procéder au préalable à une révision législative et jurisprudentielle substantielle de la notion d'« employeur » et d'« employé »⁵⁵⁵.

5.4.3. Un robot est-il l'auteur ?

L'une des options possibles - même si elle est pour le moins rémanente selon les experts⁵⁵⁶ - consisterait à considérer l'IA comme étant l'auteur. Il faudrait créer une fiction dans la loi pour doter l'IA d'une personnalité juridique, afin qu'elle puisse être l'auteur d'une œuvre et détenir un droit d'auteur⁵⁵⁷, ou tout au moins modifier la loi pour qu'elle intègre le fait qu'un ordinateur puisse être un auteur dans une œuvre commune avec une personne⁵⁵⁸. Selon R. Pearlman, la loi devrait reconnaître les IA suffisamment créatives comme des auteurs, lorsque la création de l'IA est originale et qu'elle a été développée indépendamment de toute instruction humaine, de sorte que l'IA est à l'origine de la créativité et ne se réduit pas à une simple machine travaillant selon les instructions d'un auteur humain⁵⁵⁹. Dès lors que l'IA serait déclarée auteur, les droits seraient immédiatement cédés à une personne physique ou morale, telle que le créateur/programmeur de l'IA, l'utilisateur de l'IA, ou une entité conjointe. La loi devrait identifier la personne habilitée à recevoir le transfert des droits et à les exercer.

Cependant, même dans ce scénario, l'obligation de répondre à l'exigence d'originalité pourrait être une tâche insurmontable pour une machine. Il serait nécessaire de remanier la notion d'originalité pour y inclure les œuvres provenant d'une machine, comme nous l'avons mentionné précédemment⁵⁶⁰. Cependant, si la personnalité juridique est accordée à une machine, on peut également faire valoir qu'une fois reconnue comme une personne (juridique), la machine serait capable d'une créativité originale selon l'approche personnaliste qui régit le droit d'auteur et qui conçoit l'originalité comme une expression de la personnalité de l'auteur. En tout état de cause, le fait de permettre à l'IA d'être auteur imposerait des modifications substantielles du cadre juridique. Comme nous l'avons déjà souligné, étant donné le stade précoce des développements technologiques, il

⁵⁵⁴ Bonadio E. et McDonagh L., *op.cit.*, pp. 114-115. Bridy A. (2012), *op.cit.*, p. 27, paragraphe 68. Butler T., *op.cit.*, pp. 739-742. Huson G., *op.cit.*, pp. 73-75. Ramalho A., *op.cit.*, pp. 18-19.

⁵⁵⁵ Hristov K., *op.cit.*, pp. 445-447.

⁵⁵⁶ Par exemple, Bonadio E. et McDonagh L., *op.cit.*, p. 116. Farr E.H., *op.cit.*, p. 79. Ralston W.T., *op.cit.*, pp. 302-303. Samuelson P. (1986), *op.cit.*, pp. 1199-1200.

⁵⁵⁷ Ihalainen J., « Computer creativity : artificial intelligence and copyright », *Journal of Intellectual Property Law & Practice* 13(9), pp. 724-728.

⁵⁵⁸ Abbott R., p. *op.cit.*

⁵⁵⁹ Pearlman R., *op.cit.*, p. 1-38.

⁵⁶⁰ Yanisky-Ravid S. et Velez-Hernandez L.A., *op.cit.*, pp. 40-48.

serait inopportun de modifier la loi avant même que l'on puisse construire des machines véritablement intelligentes, dont la matérialisation et l'évolution restent à ce jour une simple spéculation hypothétique⁵⁶¹.

5.4.4. Une protection *sui generis* pour la créativité artificielle

Certaines propositions ont également été faites en faveur de la création d'un droit voisin *sui generis* (sans exigence préalable en matière de paternité ou d'originalité) qui pourrait protéger les investissements consentis dans le développement et l'apprentissage de la créativité de l'IA⁵⁶². J. McCutcheon, par exemple, suggère la mise en place d'un régime de droits *sui generis* pour la créativité artificielle similaire au régime de protection des bases de données, ce qui permettrait de protéger l'investissement dans la création sans aucune exigence en matière d'auteur, de paternité ou d'originalité⁵⁶³. Tout en réfutant la protection par le régime traditionnel du droit d'auteur, l'*Australian Copyright Law Review Committee* (comité australien de révision de la législation sur le droit d'auteur) fait remarquer que si la créativité générée par ordinateur doit être protégée, cette protection devrait « s'apparenter davantage à celle qui est étendue aux droits voisins [...], celle qui est accordée aux artistes interprètes et exécutants, aux producteurs de phonogrammes et aux organismes de radiodiffusion⁵⁶⁴. »

Le Japon a envisagé un nouveau régime *sui generis* pour la propriété intellectuelle non humaine, basé sur une approche similaire au régime des marques commerciales et axé sur la protection contre la concurrence déloyale⁵⁶⁵. Cette approche vise à limiter la protection des œuvres artificielles en permettant une certaine souplesse grâce à différents niveaux de protection selon la popularité des œuvres artificielles, qui ferait office d'indicateur de valeur⁵⁶⁶. Cela écarterait les œuvres insignifiantes créées dans le seul but d'obtenir une protection au titre du droit d'auteur. Selon cette proposition, la paternité de l'œuvre serait attribuée à la personne ou à la société ayant créé l'IA⁵⁶⁷. Dans le même esprit, en vue de limiter une protection trop large de la créativité algorithmique, certains auteurs proposent un champ d'application restreint du droit *sui generis*, assorti de solides garanties d'une utilisation équitable et avec une durée courte d'environ trois ans⁵⁶⁸.

⁵⁶¹ Huson G., *op.cit.*, pp. 77-78. Liebesman Y., « The Wisdom of Legislating for Anticipated Technological Advancements », *J. Marshall Rev. Intell. Prop. L.* 10, p. 172.

⁵⁶² Ciani J., « Learning from Monkeys: Authorship Issues Arising From AI Technology » dans Moura Oliveira P., Novais P., Reis L. (éd.), *Progress in Artificial Intelligence EPIA 2019, Lecture Notes in Computer Science* vol. 11804, Springer, Cham.

⁵⁶³ McCutcheon J., *op.cit.*, pp. 965-966.

⁵⁶⁴ Ricketson S. (2012), « The Need for Human Authorship - Australian Developments: Telstra Corp Ltd v Phone Directories Co Pty Ltd (case comment) », *E.I.P.R.* 34-1, 54.

⁵⁶⁵ Intellectual Property Strategic Program 2016, pp. 10-11, http://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20160509_e.pdf.

⁵⁶⁶ Intellectual Property Strategic Program 2016, *op.cit.*, p. 11.

⁵⁶⁷ Intellectual Property Strategic Program 2016, *op.cit.*

⁵⁶⁸ Bonadio E. et McDonagh L., *op.cit.*, 136-137.

5.4.5. Accorder des droits aux éditeurs et aux diffuseurs

Enfin, d'autres propositions suggèrent d'accorder des droits aux éditeurs et aux diffuseurs des œuvres créées par l'IA. D'une part, le régime des œuvres anonymes/pseudonymes pourrait être appliqué aux œuvres créées par l'IA. En vertu de plusieurs régimes nationaux, tels qu'en Espagne, France, Italie et Suède⁵⁶⁹, la personne qui publie l'œuvre jouit de certains droits. D'autre part, une option politique supplémentaire pourrait accorder au diffuseur des créations de l'IA un droit similaire aux droits de l'éditeur au sein de l'UE concernant les œuvres non publiées auparavant, comme le prévoit l'article 4 de la Directive 2006/116/CE⁵⁷⁰. Dans le cadre de ce régime, la protection couvre la première publication/communication licite d'œuvres relevant du domaine public non publiées auparavant. De même, les œuvres créées par l'IA relèveraient du domaine public et par conséquent, le régime du « diffuseur » ne rétribuerait que la diffusion d'une création de l'IA. La durée du droit pourrait être limitée à 25 ans par exemple, comme le prévoit l'article 4 de la Directive 2006/116/CE⁵⁷¹.

5.5. Conclusion

L'anthropocentrisme et l'approche fondée sur le droit de la personnalité influencent fortement le cadre juridique actuel du droit d'auteur et, en particulier, la notion d'originalité, de la Convention de Berne jusqu'aux grandes juridictions. De ce fait, la créativité de l'IA ne satisfait pas à tous les prérequis fondamentaux pour l'octroi d'une protection au titre du droit d'auteur, notamment en ce qui concerne la personnalité juridique, la paternité et l'originalité.

Les approches utilitaristes/incitatives poussent à l'adoption de fictions juridiques qui ne sont pas en mesure de traiter ou de pallier de manière satisfaisante les incohérences systémiques internes. Néanmoins, même si la loi prétend de manière fictive que l'œuvre est réalisée par l'humain et non par l'IA, l'œuvre elle-même reste dénuée d'originalité, car les machines sont intrinsèquement incapables d'originalité selon un critère lié à la personnalité. Seule une refonte fondamentale du système des droits d'auteur mettant de côté l'approche anthropocentrique actuelle pourrait offrir une protection intégrale des droits d'auteur à la créativité de l'IA en tant que telle, en l'absence de toute intervention humaine pouvant être considérée comme une expression originale. Cette mesure serait pour le moins inconsidérée, surtout au vu du stade primitif du développement technologique dans ce domaine. Des approches *sui generis* rémanentes sont également envisageables et constituent probablement la meilleure option si les décideurs politiques

⁵⁶⁹ The Intellectual Property Act 1996, article 6, disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/es/es177en.pdf> (Espagne) ; Code de la propriété intellectuelle 1912, L113-6, disponible sur : <https://wipo.lex.wipo.int/fr/text/537285> (France) ; Law for the Protection of Copyright and Neighbouring Rights 1941, article 9, disponible sur : <https://www.wipo.int/edocs/lexdocs/laws/en/it/it211en.pdf> (Italie) ; Act on Copyright in Literary and Artistic Works 1960, article 7, disponible sur : [wipo.int/edocs/lexdocs/laws/en/se/se124en.pdf](https://www.wipo.int/edocs/lexdocs/laws/en/se/se124en.pdf) (Suède).

⁵⁷⁰ Ramalho A., *op.cit.*, pp. 22-24.

⁵⁷¹ Ramalho A., *op.cit.*

choisissent de mettre en place une incitation monopolistique en faveur de la créativité de l'IA. En ce cas, l'incitation devrait cibler les utilisateurs dès lors qu'ils contribuent par un travail et un effort significatifs au résultat de l'IA, car s'il en était autrement, les programmeurs, les distributeurs et les investisseurs profiteraient doublement des rétributions prévues en amont en lien avec le logiciel source de l'IA.

Mais des incitations sont-elles vraiment nécessaires ? La nécessité de ces incitations doit être démontrée de façon empirique, de même que les répercussions positives qu'elles pourraient avoir sur l'écosystème créatif dans son ensemble. En fait, au regard de l'histoire, on a pu établir de façon formelle que les droits de propriété n'étaient pas la seule incitation à la créativité⁵⁷². Diverses recherches et données du marché montrent que l'accès libre et gratuit aux œuvres créatives ou à des modèles commerciaux alternatifs peut inciter davantage à promouvoir la créativité artificielle que les modèles de protection basés sur la propriété intellectuelle⁵⁷³, et ce sans avoir les effets négatifs de l'appropriation et de l'exclusivité des droits. Les capacités créatives de l'IA pourraient s'accroître à un rythme étonnant, saturant le marché culturel d'un amas inextricable de droits à acquérir et d'un encombrement de droits d'auteur jusqu'à provoquer un ultime cataclysme informatique. Comme cela a déjà été dit, la soupe du droit d'auteur est déjà trop épaisse. Un nombre infini de singes IA⁵⁷⁴ peut n'en faire qu'une bouchée, ensuite il ne restera plus rien à utiliser.

⁵⁷² Par ex., Frosio G., *op.cit.*

⁵⁷³ Par ex., Bonadio E. et McDonagh L., *op.cit.*, 122-123. Svedman M., *op.cit.*, pp. 13-14.

⁵⁷⁴ Borel É., « Mécanique statistique et irréversibilité », *J. Phys. (Paris)* 5(3), pp. 189–196.

L'intelligence artificielle dans la publicité

*Les inquiétudes suscitées par l'IA sont dans une certaine mesure nourries par la littérature et les films de science-fiction. Et pourtant, la science-fiction tend à anticiper l'avenir. Prenez par exemple la scène du film « Minority Report » de Steven Spielberg⁵⁷⁵ dans laquelle un homme marche le long de l'allée d'un centre commercial. Ses yeux sont scannés par une multitude de caméras équipées de logiciels de reconnaissance des yeux. Immédiatement, les vitrines des magasins s'illuminent d'écrans criards proposant des publicités conçues sur mesure pour lui. Est-ce vraiment là de la science-fiction ? Ce n'est pas si loin de ce que vous vivons déjà aujourd'hui. À l'ère de l'internet, des téléviseurs connectés et des « deuxièmes écrans », les possibilités d'obtenir les données personnelles des utilisateurs des médias de façon légale ou illégale se sont multipliées de façon exponentielle. Ces données, qui peuvent être utilisées de façon à proposer des publicités individuellement ciblées sur les services en ligne et toutes sortes de terminaux connectés, ont une valeur capitale pour les annonceurs. De plus, les données personnelles obtenues à partir des moteurs de recherche, des médias sociaux et des appareils connectés peuvent aussi être exploitées afin d'améliorer l'expérience des utilisateurs de services en ligne. Dans sa contribution à cette publication, **Justina Raižytė** rappelle que « la création d'un cadre durable de l'intelligence artificielle en matière de publicité et l'utilisation des données et des nouvelles technologies à bon escient sont dans l'intérêt de l'industrie publicitaire elle-même », puisque c'est là le seul moyen de gagner la confiance des consommateurs, ce qui « est et restera toujours un élément incontournable en matière de publicité ».*

⁵⁷⁵ <https://www.imdb.com/title/tt0181689/>.

6. L'intelligence artificielle dans la publicité est-elle en train de devenir un univers sans foi ni loi ou les données sont-elles utilisées à bon escient?

Justina Raižytė, Alliance européenne pour l'éthique en publicité⁵⁷⁶

6.1. Introduction

L'intelligence artificielle, dont la présence se cantonnait autrefois aux scénarios des films futuristes et aux récits de fiction, a depuis un certain temps pris place dans nos salons, nos voitures, nos poches et sacs de courses. Il n'a en effet pas fallu attendre bien longtemps pour que la publicité, toujours débordante de créativité et avide d'innovation, perçoive le potentiel des technologies automatisées et fasse de l'intelligence artificielle son cheval de bataille, capable d'orienter les consommateurs dans leur vie quotidienne pour en faire de fidèles clients et utilisateurs.

Les nouvelles technologies d'apprentissage automatique permettent aux annonceurs d'analyser plus rapidement encore les données relatives aux centres d'intérêts des utilisateurs, à leurs préférences, à leur localisation et à leurs caractéristiques socioéconomiques, et même de prévoir leurs désirs, d'élaborer des stratégies de ciblage personnalisé et de leur adresser des publicités plus pertinentes. Ce processus permet de réduire le coût de la publicité, tout en augmentant son degré de réactivité et d'efficacité. Les consommateurs sont également censés être plus satisfaits et même se laisser divertir par ces publicités personnalisées, qui comportent moins de messages commerciaux ennuyeux ou irritants.

De nombreuses personnes s'inquiètent toutefois de ces flux massifs de données qui permettent aux spécialistes du marketing de segmenter les différents publics et de cibler

⁵⁷⁶ Décharge de responsabilité : les opinions exprimées dans le présent document sont celles de l'auteur et ne reflètent pas nécessairement la politique officielle ou la position de toute autre organisation, entreprise ou personne mentionnée dans le texte.

les consommateurs. Cette inquiétude et le sentiment d'être constamment traqués lors de leur navigation sur internet, d'être écoutés par des appareils électroniques dans leurs poches, ainsi que d'être sollicités par des offres dont les tarifs sont fixés de manière déloyale, amènent les consommateurs à redouter que les algorithmes puissent porter atteinte à leur vie privée et les empêcher de prendre eux-mêmes leurs décisions. C'est la raison pour laquelle l'écosystème publicitaire actuel, alimenté par la technologie, est fréquemment représenté comme un univers digne de celui des westerns, sombre, sans foi ni loi et dans lequel les principes éthiques restent fragiles.

Ce parallèle évoque inmanquablement les villes de la « ruée vers l'or », comme Deadwood, où des colons venus des quatre coins du monde espéraient faire fortune. L'histoire légendaire de cette ville a récemment pris vie sur nos écrans grâce à la magie du cinéma et à une narration impeccable qui restitue à merveille l'atmosphère chaotique du Far West, et tient lieu d'intéressante allégorie à propos du débat sur la technologie et l'évolution du marché auquel nous assistons aujourd'hui.

Où trouve-t-on l'intelligence artificielle dans la publicité d'aujourd'hui ? Quels sont les principes éthiques qui justifient l'exploitation des données à caractère personnel d'une personne et le recours à des technologies automatisées à des fins de marketing ? Quels sont les avantages et les inconvénients de l'utilisation de la puissance des algorithmes pour créer, diffuser et contrôler les publicités ? Ces questions et d'autres encore seront étudiées dans le présent article à travers le prisme des pratiques du marché et des cadres politiques en vigueur, ainsi qu'à la lumière d'entretiens qualitatifs ciblés menés avec des experts travaillant dans les domaines de l'intelligence artificielle et de la publicité, parmi lesquels des professionnels des agences publicitaires et d'entreprises technologiques, des analystes indépendants et des experts en droit de la confidentialité des données et en autorégulation de la publicité.

La première partie de l'article examinera les principales applications de l'intelligence artificielle dans la publicité et analysera la manière dont un certain nombre d'outils d'apprentissage automatique permettent de mieux cibler, d'optimiser, voire même de concevoir des communications commerciales. Le deuxième chapitre présentera les principales préoccupations et questions relatives aux applications d'apprentissage automatique dans le domaine du marketing, et examinera les garanties qui sont, ou qui devraient être, mises en place pour les atténuer. La troisième partie explorera quant à elle la manière dont les nouvelles technologies sont utilisées pour exploiter le potentiel des données au profit de bonnes causes, comme la protection des consommateurs contre les publicités frauduleuses et préjudiciables, et comment l'intelligence artificielle pourrait façonner à l'avenir le secteur de la réglementation publicitaire.

L'article aura pour conclusion la relation entre l'intelligence artificielle et la publicité à travers le prisme de la métaphore du Far West évoquée précédemment. La dernière partie résumera les arguments en examinant les similitudes et les différences entre l'époque historique de la ruée vers l'or et le phénomène de la ruée vers les données mis en avant par l'auteur, et s'interrogera sur un dernier point : sommes-nous en train de pénétrer dans un univers virtuel sans foi ni loi digne de Deadwood ?

6.2. L'intelligence artificielle dans la publicité : du traçage de la navigation en ligne à la rédaction de scénarios publicitaires

« Rien ne vaut un maximum de données ». Cette citation bien connue attribuée à Robert Mercer⁵⁷⁷ est régulièrement mentionnée pour souligner le rôle essentiel des données dans le développement et l'adoption d'algorithmes d'intelligence artificielle dans différents secteurs et activités humaines. L'utilisation de l'intelligence artificielle dans les écosystèmes publicitaires n'est pas une exception. Qu'il s'agisse de l'analyse des données à caractère personnel visant à adapter en une fraction de seconde la diffusion de publicités en fonction des préférences de l'utilisateur et de l'historique de navigation, aux algorithmes d'apprentissage destinés à reconnaître des éléments dans de vastes volumes de données passées et contextuelles (et finalement créer un nouveau contenu), l'intelligence artificielle et les techniques d'apprentissage automatique en matière de publicité sont appliquées très largement et à diverses finalités. Ce premier chapitre examinera l'utilisation principale de l'intelligence artificielle dans l'écosystème publicitaire actuel et étudiera la manière dont ces algorithmes de pointe ont transformé le secteur de la publicité.

6.2.1. La publicité programmatique : la bourse aux publicités et aux données

Selon l'*Interactive Advertising Bureau* (IAB), le succès de l'intelligence artificielle dans le secteur du commerce repose pour l'essentiel sur la publicité programmatique, c'est-à-dire le recours à l'intelligence artificielle et à d'autres technologies d'apprentissage automatique pour l'acquisition et l'optimisation d'espaces publicitaires en temps réel, afin de renforcer la rentabilité et la transparence, aussi bien pour l'annonceur que pour l'éditeur⁵⁷⁸. Dans la publicité programmatique, l'intelligence artificielle est notamment utilisée pour la planification, le placement, l'ajustement automatique, l'optimisation et la gestion des campagnes publicitaires⁵⁷⁹. Les principales caractéristiques de la publicité programmatique sont détaillées dans le « glossaire de la publicité programmatique », dans le tableau 1 ci-dessous.

⁵⁷⁷ Lee K. F., *AI Superpowers: China, Silicon Valley, and the New World Order*, Houghton Mifflin Harcourt, Boston.

⁵⁷⁸ Rask O., *What is Programmatic Advertising? The Ultimate 2020 Guide*, Match2One, <https://www.match2one.com/blog/what-is-programmatic-advertising/>.

⁵⁷⁹ IAB (2019), *Artificial Intelligence in Marketing Report*, IAB Data Center of Excellence, <https://www.iab.com/insights/iab-artificial-intelligence-in-marketing/>.

Tableau 1 Glossaire de la publicité programmatique

Enchères en temps réel (<i>Real-time bidding</i> – RTB)	
■	Les enchères en temps réel sont un moyen d'acheter et de vendre des publicités en ligne de manière ponctuelle par le biais d'enchères en temps réel qui se déroulent pendant le temps de chargement d'une page web sur le navigateur d'un internaute, c'est-à-dire environ 100 millisecondes. Parallèlement, les informations relatives à la page sur laquelle une publicité sera placée et à l'utilisateur qui charge cette page passent par une bourse de messages publicitaires, qui la propose aux enchères au profit de l'annonceur disposé à payer le prix le plus élevé, et la publicité du meilleur enchérisseur est ensuite insérée dans la page web en question ⁵⁸⁰ . Les enchères en temps réel permettent ainsi aux annonceurs de procéder à un ciblage plus efficace et plus rapide, en activant des annonces qui, par exemple, ne comportent que des publicités de la marque X pour les utilisateurs qui ont déjà visité le site internet de la marque X mais qui n'ont encore effectué aucun achat. Les annonceurs enchérissent en fonction de leurs intérêts et de la manière dont les données sur l'utilisateur correspondent à leurs critères de ciblage : plus la « correspondance » est élevée, plus le tarif est élevé ⁵⁸¹ .
Ad exchange	
■	Un <i>Ad exchange</i> est avant tout une bourse de publicité programmatique sur laquelle l'inventaire publicitaire, c'est-à-dire l'espace publicitaire, est mis aux enchères. Il s'agit d'une plateforme virtuelle où les éditeurs et les annonceurs se rencontrent, entrent en contact et conviennent d'un tarif pour diffuser leurs publicités ; toute cette procédure se déroule en quelques millisecondes, grâce à la technologie automatisée. Contrairement aux réseaux publicitaires, qui avaient pour habitude de se concentrer sur un inventaire préétabli de messages publicitaires, un <i>ad exchange</i> se fonde sur les mesures d'audience ⁵⁸² . Il est au cœur de l'écosystème programmatique, connecté aux plateformes correspondantes, tant du côté de l'annonceur que de celui de l'éditeur.
Plateforme <i>demand-side</i> (<i>Demand-side platform</i> – DSP)	
■	Une plateforme <i>Demand-side</i> est un outil ou un logiciel qui permet aux annonceurs d'acquérir automatiquement des emplacements publicitaires. Les annonceurs s'inscrivent sur une plateforme DSP, qui est à son tour connectée à un <i>ad exchange</i> ⁵⁸³ . Lorsqu'un utilisateur navigue sur un site web connecté à un <i>ad exchange</i> , un signal d'enchère est transmis. La plateforme d'échange demande ensuite à la plateforme DSP si l'annonceur dispose de publicités qui pourraient correspondre à l'emplacement et, le cas échéant, l'enchère pour cet espace publicitaire est transmise aux enchères en temps réel. Le meilleur enchérisseur peut alors diffuser sa publicité au visiteur du site concerné ⁵⁸⁴ .

⁵⁸⁰ Marshall J., *WTF is real-time bidding?*, Digiday, <https://digiday.com/media/what-is-real-time-bidding/>.

⁵⁸¹ SMAATO, *Real-Time Bidding (RTB): The Complete Guide*, SMAATO, <https://www.smaato.com/real-time-bidding/>.

⁵⁸² IAB Europe (2017), *The advent of RTB*, IAB Europe, <https://iab europe.eu/blog/lavpersons-programmatic/>.

⁵⁸³ Rask O., *op.cit.*

⁵⁸⁴ Wang J., Zhang W. et Yuan S., *Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting*, ArXiv, <https://arxiv.org/abs/1610.03013>.

Plateforme *Supply-side* (*Supply-side platform* – SSP)

- Les éditeurs utilisent quant à eux une plateforme *supply-side* (SSP) pour gérer leur espace publicitaire. La plateforme SSP se connecte à un *ad exchange* et indique précisément le type d'inventaire disponible. Grâce aux enchères en temps réel (RTB), cet inventaire est automatiquement mis aux enchères au plus offrant. Alors que l'objectif d'une plateforme DSP consiste à acquérir un espace publicitaire programmatique au meilleur prix possible auprès d'éditeurs qui disposent d'un inventaire intéressant, la plateforme SSP a une fonction inverse : vendre un espace publicitaire au prix le plus élevé possible, se connecter à plusieurs *ad exchange* différents afin d'optimiser la visibilité de l'éditeur auprès d'acheteurs potentiels⁵⁸⁵.

Bien que la publicité programmatique soit surtout utilisée par des chaînes numériques en ligne, parmi lesquelles les formats *display*, la téléphonie mobile, la vidéo et les médias sociaux, l'achat de publicités programmatiques a également trouvé sa place dans les médias traditionnels, y compris la télévision⁵⁸⁶, et même sur les panneaux d'affichage publicitaire en extérieur et la signalisation numérique⁵⁸⁷, où de puissants algorithmes associés à des données de géolocalisation mobile et à des capteurs visuels permettent le placement de publicités sur des écrans numériques aux arrêts de bus, qui peuvent désormais cibler des groupes de personnes spécifiques - les jeunes usagers aux heures de pointe du matin, par exemple. **Error! Reference source not found.**

6.2.1.1. Le rôle de l'intelligence artificielle dans la publicité programmatique

L'ensemble des éléments fondamentaux qui interviennent dans la chaîne de valeur de la publicité programmatique mettent en évidence le rôle essentiel des applications d'apprentissage automatique dans ce processus, qui permettent de prendre en un temps record des décisions en matière de sélection et de placement de messages publicitaires, rendant ainsi l'intelligence artificielle indissociable de la notion de publicité programmatique elle-même.

Ces technologies sophistiquées et automatisées permettent à tous les acteurs du marché de bénéficier de nombreux avantages basés sur la rapidité, l'efficacité et l'analyse prédictive. Dans l'ensemble, ces nombreux et vastes avantages sont résumés dans les

⁵⁸⁵ Rask O., *op.cit.*

⁵⁸⁶ Bien qu'avancées, au-delà des questions de mesure d'audience basées sur la démographie au sens large, ces questions figurent toujours parmi les principaux défis à relever pour transformer l'achat de publicités télévisées linéaires traditionnelles en programmes, certains groupes de chaînes commencent néanmoins à adopter l'achat de publicité télévisée basé sur la perception, certaines sociétés s'efforçant de normaliser les segments d'audience télévisée avancés et d'offrir un inventaire télévisé plus accessible. Voir Blustein A., *The programmatic TV dream is edging closer to reality*, The Drum, Carnyx Group Ltd.
<https://www.thedrum.com/news/2020/02/19/the-programmatic-tv-dream-edging-closer-reality>.

⁵⁸⁷ Côté R., *What is programmatic DOOH?*, Broadsign, 2020. <https://broadsign.com/blog/what-is-programmatic-digital-out-of-home/>.

objectifs fondamentaux suivants de l'intelligence artificielle dans la publicité programmatique actuelle⁵⁸⁸:

- La personnalisation des publicités, où l'intelligence artificielle et l'apprentissage automatique sont en mesure de recueillir en temps réel des données comportementales sur les consommateurs et de proposer des publicités hautement personnalisées et pertinentes, en fonction de caractéristiques telles que l'âge, le sexe, la localisation et des millions d'autres éléments d'information. Ce processus est alimenté par de puissants algorithmes prédictifs qui permettent de déterminer quels sont les consommateurs susceptibles d'interagir avec les publicités des annonceurs.
- Le ciblage de l'audience, qui utilise l'intelligence artificielle afin de permettre aux spécialistes du marketing de parcourir les contenus en ligne et de déterminer les publicités les mieux adaptées en fonction des publics ou de chaînes spécifiques. Grâce à l'utilisation de la reconnaissance d'image, ces systèmes contribuent également au placement des publicités en fonction d'images qui peuvent figurer sur la page d'un article ou d'un site web⁵⁸⁹. Par ailleurs, lorsqu'elle est correctement configurée, l'intelligence artificielle est capable de faire évoluer en permanence l'audience en fonction des véritables résultats, et de l'étendre à d'autres segments susceptibles de partager le même comportement d'achat⁵⁹⁰.
- L'optimisation des performances, où les algorithmes d'apprentissage automatique peuvent analyser automatiquement les résultats obtenus par des publicités sur des plateformes spécifiques et proposer des recommandations. Ces systèmes d'intelligence artificielle sont également capables d'analyser les résultats non seulement des campagnes publicitaires de l'annonceur, mais également de celles de ses concurrents. Ils ont une « connaissance de la situation » intégrée, qui permet aux algorithmes d'apprentissage automatique de s'ajuster rapidement, de déplacer les dépenses publicitaires vers d'autres canaux et de modifier les messages publicitaires pour refléter les modèles du marché et les comportements des consommateurs⁵⁹¹.
- La modélisation du *mix média* (mélange de médias), où l'intelligence artificielle est utilisée pour identifier les consommateurs les plus réceptifs à leurs campagnes sur différents médias, augmentant ainsi leur retour sur investissement (ROI) en matière de publicité numérique. L'intelligence artificielle peut émettre continuellement des recommandations sur la manière d'ajuster le *mix média*; les sociétés et les annonceurs peuvent donc intégralement automatiser la répartition de leur *mix marketing*, et ainsi économiser du temps et de l'argent.

En résumé, l'intelligence artificielle et l'apprentissage automatique permettent aux acteurs de l'écosystème publicitaire d'analyser d'énormes quantités de données en temps réel,

⁵⁸⁸ IAB (2019), *op.cit.*

⁵⁸⁹ Schmelzer R., « AI Makes A Splash In Advertising », *Forbes*, <https://www.forbes.com/sites/cognitiveworld/2020/06/18/ai-makes-a-splash-in-advertising/#24c0287c7682>.

⁵⁹⁰ Rowan R., « The Impact of Artificial Intelligence in Advertising », *AW360*, <https://www.advertisingweek360.com/the-impact-of-artificial-intelligence-in-advertising/>.

⁵⁹¹ Schmelzer R., *op.cit.*

d'adapter leurs messages grâce à une hyper-personnalisation de l'intelligence artificielle et ainsi déterminer les meilleurs moments et moyens pour communiquer leur message. Il n'est par conséquent pas surprenant que plus de 80 % des responsables du secteur publicitaire et des premiers utilisateurs de l'intelligence artificielle interrogés aient déclaré un retour sur investissement positif pour leurs programmes d'intelligence artificielle et exprimé leur ferme volonté d'augmenter ces investissements à l'avenir⁵⁹².

6.2.2. La créativité algorithmique : l'intelligence artificielle au service de l'imagination

« Être créatif peut s'avérer comparable au fait d'avoir une épée de Damoclès suspendue au-dessus de votre tête » – de nombreux créateurs de scénarios publicitaires qui se démènent pour respecter des délais trop courts, qui ont trop d'idées et trop peu de temps pour les concrétiser, souscriraient probablement à ces paroles d'une apparente sagesse hyperbolisée. Cette phrase n'a cependant pas été prononcée par un responsable du secteur de la publicité, mais a en réalité été générée par InspiroBot - un logiciel d'intelligence artificielle « permettant de produire des quantités illimitées de citations uniques et stimulantes », qui a été entraîné pour générer des phrases aléatoires basées sur d'innombrables citations stimulantes dénichées sur internet.

Malgré le caractère sympathique de l'exemple de « l'épée de Damoclès » généré par l'intelligence artificielle lors de la préparation de cet article, l'application réelle de l'intelligence artificielle dans la création publicitaire est bien plus sophistiquée. Les systèmes alimentés par l'intelligence artificielle sont déjà en mesure de concevoir partiellement ou intégralement des publicités, en tirant parti du traitement naturel du langage (*Natural Language Processing* – NLP) et de la génération de langage naturel (*Natural Language Generation* – NLG), deux technologies basées sur l'intelligence artificielle et utilisées pour rédiger des messages publicitaires⁵⁹³. De plus, la visualisation par ordinateur, associée aux technologies de reconnaissance d'image, annonce un avenir où les utilisateurs pourront « shazamer »⁵⁹⁴ n'importe quel contenu vidéo ou image et l'associer automatiquement aux promotions existantes.

Certaines entreprises technologiques proposent déjà des solutions basées sur l'intelligence artificielle pour générer des images à des fins commerciales. Ces entreprises, qui s'appuient généralement sur des technologies d'apprentissage automatique relèvent de la catégorie des réseaux antagonistes génératifs (*generative adversarial networks* – GAN), sont

⁵⁹² Deloitte, *State of AI in the Enterprise*, 2nd edition, the Deloitte AI Institute, Deloitte, <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/state-of-ai-and-intelligent-automation-in-business-survey.html>.

⁵⁹³ Kaput M., *AI for Advertising: Everything You Need to Know*, Marketing AI Institute, <https://www.marketingaiinstitute.com/blog/ai-in-advertising>.

⁵⁹⁴ Basé sur *Shazam Entertainment Ltd* - une application mobile qui peut identifier de la musique, des films, de la publicité et des émissions de télévision, sur la base d'un court échantillon audio joué ; d'autres types d'applications qui s'appuient sur la reconnaissance d'image ont fait l'objet de prototypes dans l'industrie, comme Smartify (disponible sur : <https://smartify.org/about-us>) – Shazam pour l'art et les musées.

capables de générer des « modèles artificiels » de plus en plus difficiles à distinguer des personnes réelles⁵⁹⁵, et peuvent ainsi créer des visages pour les campagnes publicitaires d'un professionnel du marketing, qui peuvent aisément s'adapter aux différents publics et catégories de la population.

De plus, bien que les clics de souris soient un élément essentiel du marketing numérique, l'intelligence artificielle semble faire apparaître une nouvelle tendance publicitaire : l'ère de la Voix⁵⁹⁶. Selon les données de 2019, 89 % des professionnels du secteur du marketing interrogés estimaient que les assistants vocaux constitueraient un important canal du marketing au cours des trois à cinq prochaines années, et plus d'un tiers d'entre eux les qualifiaient de « canal extrêmement important »⁵⁹⁷. Alexa, Siri, Google Home et d'autres dispositifs et assistants vocaux ont démocratisé les interfaces vocales interactives et préparé la voie à la publicité vocale interactive. Alimentés par l'intelligence artificielle et exploitant la reconnaissance vocale et la compréhension du langage naturel, ces dispositifs offrent de nouveaux formats publicitaires aux spécialistes du marketing, aux éditeurs et aux consommateurs.

Selon Charlie Cadbury, le PDG de Say It Now⁵⁹⁸, les cinq tendances les plus intéressantes en matière de voix sont : les divers fournisseurs de services permettant aux assistants vocaux d'interagir, par exemple la gestion de la livraison de produits alimentaires ; l'amélioration de la détection de compétences tierces et la reconnaissance de compétences sans nom dans les assistants vocaux ; le commerce vocal ; l'engagement audio interactif numérique ; ainsi que des services vocaux géolocalisables, disponibles dans les voitures ou les écouteurs. Interrogé sur une campagne « coup de poing » pour une application de publicité vocale, Cadbury a indiqué que « l'indice se trouve dans « l'assistant vocal » », dans lequel il espère voir les consommateurs considérer leurs assistants numériques comme un véritable atout pour leur vie quotidienne, faire entièrement confiance à leurs conseils et leur déléguer un certain nombre de tâches concrètes.

6.2.3. Des jeux créatifs aux gains

Plusieurs des principaux annonceurs à travers le globe ont déjà fait appel à des processus de création optimisés au moyen de l'intelligence artificielle, ce qui leur a permis d'obtenir des résultats et des campagnes publicitaires remarquables qui ont attiré l'attention des consommateurs et fait la une des médias. Les campagnes et les partenariats conclus entre

⁵⁹⁵ Gonfalonieri A., *Integration of Generative Adversarial Networks in Business Models*, A Medium, Towards Data Science,

<https://towardsdatascience.com/integration-of-generative-adversarial-networks-in-business-models-47e60263aec4>.

⁵⁹⁶ Tushinskiy S., *Voice is the New Click: Why Voice Commands Will Replace the Click as the Standard Measure for Brands*, Medium, Instreamatic, <https://medium.com/@instreamatic/voice-is-the-new-click-why-voice-commands-will-replace-the-click-as-the-standard-measure-for-61225e4e7caa>.

⁵⁹⁷ Kinsella B. et Mutchler A., *The State of Voice Assistants as a Marketing Channel Report*, Voicebot.ai, 2019 <https://voicebot.ai/the-state-of-voice-assistants-as-a-marketing-channel-report/>.

⁵⁹⁸ Charlie Cadbury de *Say It Now* (disponible sur : <https://www.sayitnow.ai/>) a été interviewé le 19 juin 2020 pour les besoins du présent article.

des marques et des agences publicitaires, présentés dans le tableau 2, offrent une sélection de cas de figure dans lesquels des outils de l'intelligence artificielle ont été utilisés à des fins commerciales créatives et ont abouti à la production de contenus publicitaires conçus, au moins partiellement, par des technologies automatisées. Kerry Richardson, l'un des cofondateurs de Tiny Giant⁵⁹⁹ estime que les consommateurs verront sous peu davantage d'applications créatives en matière d'intelligence artificielle, et davantage de collaboration entre des personnes physiques et des applications créatives pour l'élaboration de campagnes publicitaires, où les professionnels du secteur de la création imagineront une idée ou un concept, puis utiliseront la technologie de l'intelligence artificielle pour y parvenir.

Tableau 2 Campagnes de publicité et de marketing optimisées par les technologies créatives de l'intelligence artificielle

■ Lexus en partenariat avec Watson d'IBM, The & Partnership London et Visual Voice⁶⁰⁰
■ Cette publicité spécifique pour la marque Lexus est la première publicité télévisée dont le scénario a été entièrement conçu au moyen de l'intelligence artificielle. En utilisant un certain nombre de paramètres, dont 15 années de publicités récompensées par les Lions de Cannes et 10 années des meilleures publicités du secteur du « luxe », ainsi que des éléments relatifs à « l'intuition » et à la manière dont les individus prennent des décisions, la technologie IBM Watson a identifié des éléments communs aux publicités récompensées⁶⁰¹. Un moteur d'intelligence artificielle a ensuite élaboré la trame et les grandes lignes du scénario. Les créateurs eux-mêmes ont fait part de leur surprise sur le fait qu'au lieu du résultat « délirant et bizarre » auquel ils s'attendaient, ils ont obtenu un scénario avec une note de bas de page à chaque ligne, accompagnée d'un commentaire expliquant pourquoi telle décision avait été prise⁶⁰². Le résultat s'est traduit par une histoire plutôt déroutante d'un ingénieur Lexus mettant la touche finale à un nouveau modèle, une voiture consciente de ses propres capacités et un crash test télévisé, mais cela a démontré que la version créative conçue par une machine est bien plus qu'une éventuelle possibilité.

⁵⁹⁹ Kerry Richardson de Tiny Giant (disponible sur : <https://www.tinygiant.io/>) a été interviewé le 22 juin 2020 pour les besoins du présent article.

⁶⁰⁰ La campagne publicitaire peut être visionnée sur Variety.com. Disponible sur : <https://variety.com/2018/digital/news/lexus-ai-scripted-ad-ibm-watson-kevin-macdonald-1203030693/>.

⁶⁰¹ Spangler T., *First AI-Scripted Commercial Debuts* ; Réalisé par Kevin Macdonald pour Lexus, Variety, <https://variety.com/2018/digital/news/lexus-ai-scripted-ad-ibm-watson-kevin-macdonald-1203030693/>.

⁶⁰² Faull J., *Lexus reveals ad 'created by AI'. Is it a gimmick? No. Will it win any awards? Probably not*, The Drum, Carnyx Group Ltd, <https://www.thedrum.com/news/2018/11/16/lexus-reveals-ad-created-ai-it-gimmick-no-will-it-win-any-awards-probably-not>.

■	Malaria No More en partenariat avec RG/A, Ridley Scott Associates et Synthesia ⁶⁰³
■	Dans la campagne « <i>Malaria Must Die</i> » (« Il faut éradiquer la malaria »), David Beckham prête sa voix à la lutte contre le paludisme de l'organisation à but non lucratif Malaria No More, dans sa bataille pour éradiquer la maladie transmise par les moustiques. Cette campagne innovante présente David Beckham parlant apparemment couramment neuf langues et invitant les auditeurs à se mobiliser pour cette cause ⁶⁰⁴ . Alors qu'il passe d'une langue à l'autre, les différentes voix de David Beckham sont en fait celles de survivants du paludisme dont les caractéristiques ont été cartographiées numériquement sur celle du célèbre sportif à l'aide de la technologie de synthèse vidéo assistée par l'intelligence artificielle ⁶⁰⁵ .
■	Partenariat entre la Deutsche Bahn et Ogilvy Germany, Francfort, Getty Images et Spirable ⁶⁰⁶
■	La Deutsche Bahn, la société ferroviaire allemande, a lancé une campagne encourageant les voyages sur le territoire national en utilisant des photos de sites allemands qui évoquent des destinations touristiques mondiales très connues. La campagne « No Need To Fly » (« Pas besoin de prendre l'avion ») invite les Allemands à profiter des avantages d'un voyage en train moins cher. L'intelligence artificielle a été utilisée pour identifier des sites allemands ressemblant à des lieux touristiques emblématiques du globe. Ensuite, grâce aux données de Facebook, la campagne a ciblé les passionnés de voyages et les influenceurs locaux avec des publicités vidéo dynamiques dans lesquelles se juxtaposaient en temps réel les frais relatifs à ces sites internationaux et à leurs équivalents allemands ⁶⁰⁷ .
■	Partenariat entre Cheltenham Science Festival et Tiny Giant ⁶⁰⁸
■	« AIDA : <i>AI Science Festival Curator</i> », un organisateur de festival généré par l'intelligence artificielle pour le Festival Cheltenham Science, a rassemblé comme base de données 10 années de conférences du festival, ces données ont été intégrées sur un réseau neuronal récurrent et ont généré environ 800 nouvelles

⁶⁰³ La campagne publicitaire peut être visionnée sur TheDrum.com. Disponible sur :

<https://www.thedrum.com/news/2019/04/09/david-beckham-lends-his-voice-malaria-ai-petition>.

⁶⁰⁴ Butcher, M., The startup behind that deep-fake David Beckham video just raised \$3M, TechCrunch, <https://techcrunch.com/2019/04/25/>.

⁶⁰⁵ Glenday J., *David Beckham lends his voice to Malaria AI petition*, The Drum, Carnyx Group Ltd., <https://www.thedrum.com/news/2019/04/09/david-beckham-lends-his-voice-malaria-ai-petition>.

⁶⁰⁶ La campagne publicitaire peut être visionnée sur wersm.com. Disponible sur : <https://wersm.com/how-deutsche-bahn-increased-sales-by-24-thanks-to-instagram/>.

⁶⁰⁷ Desreumaux G., *How Deutsche Bahn Increased Sales By 24% Thanks To Instagram*, Wersm, <https://wersm.com/how-deutsche-bahn-increased-sales-by-24-thanks-to-instagram/>.

⁶⁰⁸ La campagne publicitaire peut être visionnée sur TinyGiant.io. Disponible sur : <https://www.tinygiant.io/case-study-one-aida>.

conférences potentielles pour le festival⁶⁰⁹. Les propositions d'AIDA ont fait l'objet d'un sondage sur Twitter pour que le monde entier puisse participer à la sélection d'un gagnant. La conférence « Introvert Narwhals » (« Les narvals introvertis ») a été diffusée lors de l'événement par l'AIDA, dont la voix audio a été créée à l'aide d'un apprentissage en profondeur pour transformer le texte saisi en une voix nuancée et à consonance humaine. AIDA a connu un vif succès et a ensuite fait l'objet d'une diffusion sur BBC Radio, et a également reçu le prix d'or de l'association Data & Marketing pour la meilleure utilisation de l'intelligence artificielle en 2019⁶¹⁰.

■ Partenariat permanent entre JPMorgan Chase et Persado

- En 2019, JPMorgan Chase a annoncé la conclusion d'un accord d'entreprise d'une durée de cinq ans avec Persado, l'une des principales agences dans le domaine de l'intelligence artificielle, afin de générer des créations marketing particulièrement performantes. Leur projet pilote a démontré que ces messages commerciaux générés par l'intelligence artificielle étaient très efficaces, avec une augmentation de 450 % des clics pour les publicités diffusées par Persado, par rapport à d'autres qui fluctuent entre 50 à 200 %⁶¹¹. Les technologies exclusives de Persado ont été utilisées pour réécrire les textes et les titres en s'appuyant sur une base de connaissances avancées du langage commercial de plus d'un million de mots et d'expressions identifiés et évalués. Grâce à cet outil, JPMorgan Chase a remanié les messages commerciaux dans ses activités de crédits et de prêts hypothécaires, afin de les rendre les plus convaincants possible pour les particuliers et les groupes de clients ciblés⁶¹².

6.2.3.1. Le caractère virtuel du marketing d'influence : une étude de cas

Lil Miquela est une chanteuse qui diffuse de la musique en streaming sur Spotify, une styliste propriétaire de sa marque de vêtements, un mannequin travaillant avec des grandes enseignes de vêtements de luxe et une star des médias sociaux avec plus de 2,4 millions d'abonnés sur son compte Instagram⁶¹³. Elle se définit elle-même comme « une musicienne,

⁶⁰⁹ Tiny Giant, *AI Festival curator Cheltenham Science Festival*, Tiny Giant, <https://www.tinygiant.io/case-study-one-aida>.

⁶¹⁰ Data & Marketing Association, *Gold Best Use of AI*, Data & Marketing Association Awards, <https://dma.org.uk/awards/winner/2019-gold-best-use-of-ai>.

⁶¹¹ Persado, *JPMorgan Chase Announces Five-Year Deal with Persado For AI-Powered Marketing Capabilities*, Persado, <https://www.persado.com/press-releases/jpmorgan-chase-announces-five-year-deal-with-persado-for-ai-powered-marketing-capabilities/>.

⁶¹² Business Wire, *JPMorgan Chase Announcement Concerning Preferred Stock*, Business Wire, <https://www.businesswire.com/news/home/20191031005537/en/JPMorgan-Chase-Announcement-Preferred-Stock>.

⁶¹³ Au 1^{er} juillet 2020.

une adepte du changement et un robot » – un personnage généré par ordinateur (CGI) et la première influenceuse virtuelle de notoriété mondiale.

Bien que Lil Miquela ne soit pas véritablement une création de l'intelligence artificielle, elle a incité les entreprises à investir massivement dans les êtres humains virtuels et à imaginer de futurs êtres numériques entièrement conçus au moyen de l'intelligence artificielle et présents de manière autonome sur les plateformes des médias sociaux⁶¹⁴. Comme la production de modèles 3D de haute précision devient de plus en plus abordable, un certain nombre de créateurs envisagent déjà des humains virtuels « vivant » leur propre existence sans aucune implication humaine : de la publication de photos ou de vidéos, et de commentaires qui les accompagnent, à une interaction avec leurs abonnés⁶¹⁵.

Bien que l'idée d'humains virtuels, animés par des machines, ne soit pas une nouveauté, « la prolifération des smartphones et la popularité des sites de partage d'images comme Instagram ont contribué à notre prise de conscience de l'existence de ces humains virtuels et les ont élevés à une position d'influence », déclare Scott Guthrie⁶¹⁶, consultant et analyste indépendant en marketing d'influence. L'industrie de la mode a été la première à exploiter les potentiels atouts qu'offrent les modèles virtuels : les humains synthétiques n'ont en effet pas besoin de régimes de beauté, de tailles de vêtements ajustées ou de multiples prises lors d'une séance photo ; ils présentent toujours les vêtements de la marque du sponsor de la meilleure façon qui soit, arrivent toujours à l'heure et offrent toujours un contenu conforme au cahier des charges⁶¹⁷.

6.2.3.1.1. Les véritables dangers de la réalité artificielle

Malgré leur caractère virtuel, les êtres numériques suscitent de véritables préoccupations, au premier rang desquelles figure la transparence. Rupa Shah, fondatrice de Hashtag Ad Consulting⁶¹⁸, observe que le déploiement de technologies avancées signifie qu'il devient de plus en plus difficile de distinguer les influenceurs virtuels d'un simple coup d'œil. Le souci du détail dans le rendu de chaque image leur permet d'apparaître dans n'importe quel contexte ou scène et à n'importe quel endroit, afin de réaliser le récit souhaité par la marque, et le traitement naturel du langage (NLP) confère à leurs discours un caractère naturel et vivant.

Les codes d'autorégulation de la publicité, appliqués par les organismes de régulation de la publicité à travers l'Europe, exigent déjà que toutes les communications

⁶¹⁴ Alexander J., *Virtual creators aren't AI – but AI is coming for them*, The Verge, Vox Media, <https://www.theverge.com/2019/1/30/18200509/ai-virtual-creators-lil-miquela-instagram-artificial-intelligence>.

⁶¹⁵ Bradley S., *Even better than the real thing? Meet the virtual influencers taking over your feeds*, The Drum, Carnyx Group Ltd, <https://www.thedrum.com/news/2020/03/20/even-better-the-real-thing-meet-the-virtual-influencers-taking-over-your-feeds>.

⁶¹⁶ Scott Guthrie, *consultant, analyste indépendant en marketing d'influence et blogger*. Disponible sur : <https://sabguthrie.info/>. Interviewé le 22 juin 2020 pour les besoins du présent article.

⁶¹⁷ Guthrie S., « Virtual Influencers: More Human Than Human », chapitre 15, S. Yesilogl et J. Costello, (sous la direction de), *Influencer Marketing: Building Brand Communities and Engagement*, Routledge, London.

⁶¹⁸ Rupa Shah de *Hashtag Ad Consulting*. Disponible sur : <https://www.hashtagad.co.uk/>. Interviewé le 22 juin 2020 pour les besoins du présent article.

commerciales soient identifiables immédiatement et sans ambiguïté, au moyen d'une information appropriée⁶¹⁹. Ces dispositions devront sans doute être étendues aux influenceurs virtuels, en imposant par ailleurs à leurs propriétaires et créateurs l'obligation d'informer les consommateurs de la nature virtuelle de ces influenceurs, qui sont intégralement contrôlés par une marque.

Dudley Neville-Spencer, directeur et responsable de l'analyse des données de la *Virtual Influencer Agency*⁶²⁰, reconnaît que les influenceurs virtuels sont susceptibles d'amplifier les problèmes déjà existants dans le marketing d'influence et plaide pour un marquage spécifique des êtres virtuels⁶²¹. Il collabore avec Shah à l'élaboration d'un code d'éthique des influenceurs virtuels, en mettant non seulement l'accent sur la transparence, mais également sur d'autres aspects de la responsabilité sociale, tels que l'image corporelle et la diversité. De nombreuses voix réputées du secteur ont exprimé leur inquiétude sur le fait que les influenceurs virtuels pourraient entraîner des problèmes d'estime de soi et de santé psychique, en raison d'une représentation humaine préjudiciable et de critères de beauté idéalisés et irréalistes⁶²². Shah ajoute qu'il importe que les régulateurs se préparent à cette situation et qu'ils soient prêts à répondre à l'ensemble des craintes des consommateurs.

6.2.3.1.2. L'avenir des influenceurs virtuels

À l'échelle mondiale, le marché des influenceurs progresse rapidement et devrait représenter 9,7 milliards de dollars américains en 2020⁶²³. L'intelligence artificielle n'est pas seulement appliquée pour créer des influenceurs virtuels, elle est également employée dans les phases de sélection et d'évaluation du marketing d'influence, aidant ainsi les marques à identifier les influenceurs des médias sociaux les plus pertinents pour d'éventuelles campagnes publicitaires et à les examiner (par exemple, en vérifiant le nombre d'abonnés potentiellement factices).

Guthrie estime que de nouvelles évolutions technologiques permettront aux influenceurs virtuels de « se libérer » des modèles d'animation préétablis, d'interagir librement et de tirer parti de chaque conversation humaine. Il considère également qu'à l'avenir, le sous-genre des influenceurs virtuels se subdivisera en au moins trois catégories plus petites : les assistants virtuels d'une marque (c'est-à-dire des êtres virtuels conçus et gérés par une marque, et entièrement dévoués aux objectifs de la marque en question) ;

⁶¹⁹ Alliance européenne pour l'éthique en publicité, *Best Practice Recommendation on Influencer Marketing*, EASA, disponible en anglais sur :

<https://www.easa-alliance.org/products-services/publications/easa-best-practice-guidance>.

⁶²⁰ Les commentaires de Dudley Neville-Spencer, directeur et responsable de l'analyse des données de la Virtual Influencer Agency et directeur de la stratégie et des idées à Live & Breath, sont extraits d'un webinaire en ligne organisé par Persollo le 30 juin 2020, voir Persollo, *Ethics, Influencers and Growth*, Persollo Webinar 3, 30 juin 2020,

<https://www.blog.persollo.com/post/persollo-webinar-3-ethics-influencers-and-growth>.

⁶²¹ Persollo, *op.cit.*

⁶²² Bradley S., *op.cit.*

⁶²³ Influencer Marketing Hub, *The State of Influencer Marketing 2020: Benchmark Report*, Influencer Marketing Hub, <https://influencermarketinghub.com/influencer-marketing-benchmark-report-2020/>.

des représentants du service clientèle (c'est-à-dire des êtres virtuels fonctionnant comme des agents conversationnels (*chatbots*), mais avec des émotions et une apparence corporelle semblables à celles d'un être humain); et des influenceurs virtuels, à savoir des influenceurs humains virtuels appartenant à la marque d'un sponsor ou opérant de leur propre chef.

Enfin, dans la mesure où les influenceurs virtuels continuent à évoluer, il reviendra au public de décider de la fréquence à laquelle il sera prêt à adopter des avatars numériques dans son fil d'actualités sur les médias sociaux et à interagir avec eux. La conviction de certains experts, selon laquelle toutes les marques disposeront d'une sorte de représentant virtuel, deviendra peut-être réalité dans un futur proche⁶²⁴.

6.2.4. Conclusion: la publicité intelligente grâce à l'intelligence artificielle

La publicité intelligente est réputée être une troisième phase et une nouvelle ère de la publicité numérique, qui repose sur le marketing interactif et programmatique⁶²⁵. Elle exploite l'interactivité et l'automatisation des phases précédentes, mais intègre de nouveaux éléments, tels que la personnalisation qui ne se limite plus aux préférences de l'utilisateur, et s'oriente vers la prédiction de ses besoins dans divers contextes et moments, en diffusant en temps réel et à grande échelle des contenus commerciaux extrêmement personnalisés⁶²⁶. Elle fusionne les techniques d'achat et de créativité programmatiques et ouvre la voie à des expériences publicitaires sur mesure qui devraient permettre d'obtenir un contenu encore plus pertinent et utile.

Toutefois, pour que la publicité intelligente soit efficace et gagne la confiance des consommateurs, elle ne peut pas être enfermée dans une boîte noire opaque qui génère des messages promotionnels basés sur des données d'utilisateur collectées de manière illicite et des algorithmes inappropriés. La suite du présent chapitre se penchera donc les défis que pose l'intelligence artificielle dans la publicité et examinera les garanties qu'il serait nécessaire de mettre en place.

6.3. Les inquiétudes concernant les données massives (Big Data) et l'intelligence artificielle

Aujourd'hui, la plupart des utilisations de l'intelligence artificielle dans la publicité reposent sur des algorithmes et de grands ensembles de données qui contiennent des informations

⁶²⁴ Persollo, *op.cit.*

⁶²⁵ Li H., « Special Section Introduction: Artificial Intelligence and Advertising », *Journal of Advertising*, août/septembre 2019, Volume 48, n° 4, pp. 333-337.

⁶²⁶ Chen G. et al., « Understanding Programmatic Creative: The Role of AI », *Journal of Advertising*, août/septembre 2019, Volume 48, n° 4, pp. 347-355.

sur les habitudes et les préférences de chaque utilisateur, permettant ainsi de diffuser un contenu commercial personnalisé et « soigneusement sélectionné »⁶²⁷. Un tel processus d'optimisation pose bien entendu un certain nombre de défis éthiques et juridiques, notamment en matière de confidentialité et d'algorithmes biaisés. Ce chapitre examinera donc les garanties juridiques existantes qui visent à protéger les citoyens et les consommateurs contre tout éventuel préjudice en matière de protection des données et de prise de décision automatisée. Bien que cette partie se limite au cadre européen, il convient de noter qu'il existe ailleurs dans le monde d'autres politiques relatives à la confidentialité et à la protection des données qui pourraient également faire l'objet d'une analyse similaire⁶²⁸.

Les experts ont estimé qu'une législation efficace de protection de la vie privée à l'ère de l'intelligence artificielle devrait protéger les consommateurs contre la discrimination, l'absence de consentement et l'utilisation illicite des données obtenues par le recours à l'intelligence artificielle, et qu'elle devrait comporter plusieurs éléments essentiels, à savoir une exigence de transparence, afin que l'intelligence artificielle puisse disposer d'un véritable droit sur les informations qu'elle recueille ; une option de refus pour les consommateurs ; les données collectées et l'objectif de la limitation de l'intelligence artificielle par finalité ; ainsi que la possibilité d'une suppression des données à la demande du consommateur⁶²⁹. Les groupes de la société civile ont également demandé « des lignes rouges précises pour les utilisations non autorisées, un contrôle démocratique et une stratégie de réglementation de l'intelligence artificielle véritablement axée sur les droits fondamentaux », de manière à élaborer un système d'intelligence artificielle parfaitement fiable⁶³⁰. Bien que le présent document se limite à l'utilisation de la technologie automatisée dans le secteur du marketing, il importe également de tenir compte de la question des droits de l'homme, et tout particulièrement du droit au respect de la vie privée et à la protection des données (article 8, Convention européenne des droits de l'homme) et de l'interdiction de la discrimination (article 14 de la même Convention), et notamment la manière dont les données à caractère personnel sont collectées, les modalités de leur réutilisation et comment des algorithmes conçus avec des biais intentionnels et involontaires pourraient conduire à une segmentation et à un traitement différencié des

⁶²⁷ Lee K. F., *op.cit.*

⁶²⁸ Par exemple, la California Consumer Privacy Act (CCPA), qui prévoit de nouveaux droits des consommateurs au sujet de l'utilisation, la suppression, le retrait et l'accès des entreprises aux informations à caractère personnel (voir A. Paka, *How Does The CCPA Impact Your AI?*, Forbes Technology Council, Forbes, <https://www.forbes.com/sites/forbestechcouncil/2020/02/20/how-does-the-ccpa-impact-your-ai/#7d27ce7c43c7>), ou la loi relative à la protection des informations à caractère personnel au Japon, qui devrait être à nouveau modifiée en 2020 avec l'insertion et la promotion de données pseudonymisées pour alimenter l'intelligence artificielle, voir GLI, *AI, Machine Learning & Big Data Laws and Regulations Japan*, Global Legal Insights, <https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/japan>. Voir également le chapitre 2 de la présente publication.

⁶²⁹ Intel AI, *Rethinking Privacy For The AI Era*, Forbes, Insight team, <https://www.forbes.com/sites/insights-intelai/2019/03/27/rethinking-privacy-for-the-ai-era/#693cda737f0a>.

⁶³⁰ EDRI, *Can the EU make AI "trustworthy"? No – but they can make it just*, EDRI, <https://edri.org/can-the-eu-make-ai-trustworthy-no-but-they-can-make-it-just/>.

groupes sociaux (ou de consommateurs) ciblés⁶³¹. Enfin, les experts ont également mis en garde contre « le déséquilibre de pouvoir entre ceux qui développent et appliquent les technologies de l'intelligence artificielle et ceux qui interagissent avec et y sont soumis »⁶³². Compte tenu de ces problèmes, la partie suivante examinera les mécanismes existants en matière de responsabilité et d'obligation de rendre des comptes afin de faire face à ces risques et de les atténuer.

6.3.1. Le cadre juridique en vigueur en Europe

En février 2020, la Commission européenne a publié un Livre blanc proposant une stratégie pour assurer le succès de l'adoption de l'intelligence artificielle dans l'Union européenne au moyen d'un cadre politique et réglementaire approprié et de la création d'un « écosystème d'excellence et de confiance »⁶³³. Le Livre blanc rappelle que l'Union européenne s'est dotée d'un cadre juridique strict pour assurer, notamment, la protection des consommateurs, lutter contre les pratiques commerciales déloyales et assurer la protection des données à caractère personnel et de la vie privée, notamment le « Règlement général relatif à la protection des données » et d'autres dispositions sectorielles applicables à la protection des données à caractère personnel telles que la « Directive relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel »⁶³⁴. Bien que le RGPD ne mentionne pas spécifiquement l'intelligence artificielle, il est structuré de manière technologiquement neutre, afin de faire face à tout changement ou évolution technologique, et s'applique donc pleinement au traitement des données à caractère personnel par l'intermédiaire d'un algorithme⁶³⁵. Par ailleurs, même les systèmes d'intelligence artificielle qui exploitent des données anonymisées peuvent être soumis aux dispositions du RGPD, dans la mesure où certaines techniques d'anonymisation ne permettent pas toujours d'annuler le risque d'une ré-identification⁶³⁶. En outre, selon une évaluation du comité européen de la protection des données (CEPD), une « stratégie basée sur les risques, le principe de minimisation des données et l'exigence de protection des données dès la conception et par défaut », ainsi que des dispositions en matière de responsabilité, la transparence et l'interdiction de toute

⁶³¹ Wagner B., *Étude sur les dimensions des droits humains dans les techniques de traitement automatisé des données (en particulier les algorithmes) et éventuelles implications réglementaires*, Conseil de l'Europe, Comité d'expert sur les intermédiaires d'internet (MSI-NET), <https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a>.

⁶³² Yeun K., *Responsabilité et IA*, Étude du Conseil de l'Europe DGI(2019)05, Conseil de l'Europe, Comité d'experts sur les dimensions des droits de l'homme dans le traitement des données et des différentes formes d'intelligence artificielle (MSI-AUT), <https://rm.coe.int/responsability-and-ai-fr/168097d9c6>.

⁶³³ Commission européenne, *Livre blanc – Intelligence artificielle – Une approche européenne axée sur l'excellence et la confiance*, Bruxelles, 19 février 2020, COM(2020) 65, https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

⁶³⁴ Commission européenne, *op.cit.*

⁶³⁵ Comité européen de la protection des données, *EDPB Response to the MEP Sophie in 't Veld's letter on unfair algorithms*, EDPB, disponible en anglais sur : https://edpb.europa.eu/our-work-tools/our-documents/letters/edpb-response-mep-sophie-int-velds-letter-unfair-algorithms_en.

⁶³⁶ Zaccaria E., *Artificial Intelligence: addressing the risks to data privacy and beyond*, PrivSec Report, <https://gdpr.report/news/2020/06/01/artificial-intelligence-addressing-the-risks-to-data-privacy-and-beyond/>.

prise de décision reposant uniquement sur un traitement automatisé, permettraient d'adapter le cadre juridique en vigueur pour faire face aux nombreux risques et défis potentiels relatifs au traitement des données à caractère personnel par des algorithmes⁶³⁷. Enfin, un rapport consacré à l'impact du RGPD sur l'intelligence artificielle, publié en juin 2020 par le Service de recherche du Parlement européen (EPRS), conclut que le RGPD « fournit des indications importantes pour la protection des données dans le contexte des applications d'intelligence artificielle », et qu'il convient que les « responsables du traitement souscrivent aux valeurs du RGPD et adoptent une stratégie responsable et axée sur les risques »⁶³⁸.

Géraldine Proust, directrice des questions politiques à la Fédération européenne du marketing direct et interactif (FEDMA)⁶³⁹, qui a fait part de son expertise sur les lois relatives à la protection des données dans l'Union européenne, partage cet avis et affirme que le RGPD comporte en effet des principes qui protègent les données à caractère personnel dont le traitement est effectué par des technologies utilisant l'intelligence artificielle, et qu'il exige que les entreprises rendent des comptes, qu'elles ne se contentent pas uniquement de respecter ces principes et qu'elles soient également en mesure de démontrer qu'elles les respectent. Elle précise par ailleurs que « les organisations concernées ne doivent traiter que les données à caractère personnel qui s'avèrent nécessaires, pertinentes et vérifiables » et que « le traitement doit en outre être équitable, transparent, conforme à la loi et poursuivre un objectif légitime. Si cette finalité peut être obtenue sans l'utilisation de données à caractère personnel, une autre méthode doit être envisagée, par exemple l'anonymisation des données ». Cela suppose que les experts en technologies de l'information et les gestionnaires et les responsables chargés de la protection des données travaillent ensemble le plus tôt possible dans le processus de création marketing, afin de parvenir à un juste équilibre entre innovation et éthique. Le comité européen de la protection des données a également indiqué que, compte tenu du cadre juridique en vigueur déjà très complet, il convient de mettre l'accent sur le renforcement des normes existantes, sur l'obligation de rendre des comptes et sur les analyses d'impact des mesures de protection des données des algorithmes d'apprentissage automatique. Géraldine Proust souscrit à cette initiative et estime qu'il importe tout d'abord que la législation en vigueur soit correctement mise en œuvre et évaluée, et que toute nouvelle législation « doit être équilibrée pour éviter d'entraver le développement et l'utilisation de cette technologie et être conforme à la législation applicable à la protection des données afin de prévenir toute contradiction ».

Les lignes directrices publiées par les autorités compétentes pourraient être un moyen efficace d'apporter des précisions et des conseils supplémentaires sur l'application d'exigences juridiques spécifiques, le cas échéant. Un récent rapport du Service de recherche du Parlement européen (EPRS) met l'accent sur la nécessité de ces orientations

⁶³⁷ Comité européen de protection des données, *op.cit.*

⁶³⁸ Service de recherche du Parlement européen (2020), *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*, Étude du Service de recherche du Parlement européen, Unité de prospective scientifique (STOA), PE 641.530, disponible en anglais sur : [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf).

⁶³⁹ Géraldine Proust de la FEDMA. Disponible sur : <https://www.fedma.org/>. Interviewée le 7 juillet 2020 pour les besoins du présent article.

et rappelle qu'il importe que les responsables du traitement et les personnes concernées « ne soient pas abandonnés » et « puissent être informés sur la manière dont l'intelligence artificielle s'applique à leurs données personnelles, conformément au RGPD »⁶⁴⁰. Les auteurs du rapport préconisent une action à plusieurs niveaux et invitent les institutions à s'engager activement dans de vastes débats de société avec l'ensemble des parties prenantes, y compris les responsables du traitement, les sous-traitants et la société civile, afin d'élaborer des réponses appropriées et des suggestions pertinentes « fondées sur des valeurs communes et des technologies efficaces »⁶⁴¹. À cette fin, plusieurs éléments d'orientation intéressants ont déjà été publiés par les institutions européennes, y compris la Commission européenne, le Conseil de l'Europe et le comité européen de la protection des données⁶⁴², ainsi que par des professionnels du secteur⁶⁴³. De nombreuses autres recommandations sont encore susceptibles d'être adoptées au cours des prochaines années et, accompagnées d'instruments de régulation et d'autorégulation, elles constitueront, espérons-le, de précieux outils en faveur du succès global de l'adoption de l'intelligence artificielle, de la protection des consommateurs et de leur confiance dans les nouvelles technologies.

6.3.2. Conclusion : (essentiellement) le Bon, la Brute et le Truand

La réponse aux principales inquiétudes que suscite aujourd'hui l'utilisation de l'intelligence artificielle dans l'univers du marketing s'inscrit parfaitement dans l'esprit du célèbre western de Sergio Leone « Le Bon, la Brute et le Truand ».

Dans l'ensemble, les nouvelles sont plutôt bonnes, puisque la législation et les orientations politiques existantes aujourd'hui offrent un cadre bien défini pour mettre en place, dans le domaine de la publicité, des systèmes d'intelligence artificielle pérennes qui protègent les droits de l'homme et les intérêts des consommateurs sans entraver l'innovation. Par ailleurs, les entreprises, depuis les multinationales technologiques jusqu'aux start-ups, semblent avoir réévalué les avantages à long terme de la confiance des

⁶⁴⁰ Service de recherche du Parlement européen, *op.cit.*

⁶⁴¹ Service de recherche du Parlement européen, *op.cit.*

⁶⁴² Par exemple, les Lignes directrices sur l'intelligence artificielle et la protection des données (2019) du Comité consultatif de la Convention pour la protection des personnes à l'égard du traitement des données à caractère personnel (Convention 108). Disponibles sur : <https://rm.coe.int/lignes-directrices-sur-l-intelligence-artificielle-et-la-protection-de/168091ff40>. Conseil de l'Europe, Lignes directrices en matière d'éthique pour une IA digne de confiance (2019), établies par le Groupe d'experts de haut niveau sur l'intelligence artificielle, Commission européenne. Disponibles sur : <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>; Lignes directrices 4/2019 sur l'article 25 de la protection des données par conception et par défaut (2019), établies par le comité européen de la protection des données. Disponible sur : https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design_en.

⁶⁴³ Par exemple, High Level Principles on Artificial Intelligence (2020) de la European Tech Alliance (EUTA). Disponible sur <http://eutechalliance.eu/wp-content/uploads/2020/02/EUTA-High-Level-Principles-on-AI.pdf>; The Seven-Step Ad Tech Guide (2020) de la Data & Marketing Association (DMA) et Incorporated Society of British Advertisers (ISBA) addressing privacy challenges of RTB in programmatic advertising. Disponible sur : <https://dma.org.uk/uploads/misc/seven-step-ad-tech-guide-v9.pdf>.

consommateurs et appellent à une démarche davantage au service de l'humain lors de la conception et de l'application de l'intelligence artificielle, depuis ses capacités de fonctionnement, de détection, de cognition et d'apprentissage⁶⁴⁴, jusqu'aux conséquences sur la manière dont elle influence les utilisateurs finaux (approche micro-économique), ainsi que la société et les principes d'éthique en général (approche macro-économique). La société néerlandaise DEUS qualifie cette démarche de marketing d'approche « centrée sur l'humanité », où elle est particulièrement pertinente, puisque la publicité répond aux besoins, aux désirs et, bien souvent, aux tentations d'une personne dans son processus d'achat. Comme les consommateurs prennent de plus en plus conscience de l'utilisation de leurs données à des fins publicitaires et qu'ils souhaitent en savoir davantage encore⁶⁴⁵, il semble inévitable que la question de l'utilisation éthique de l'intelligence artificielle dans le marketing continue à être une priorité pour la société civile, les décideurs politiques et les professionnels sérieux du secteur, qui s'accordent apparemment tous sur la nécessité d'un système d'intelligence artificielle digne de confiance.

La mauvaise nouvelle tient quant à elle au fait que l'ensemble des politiques et stratégies précitées ne seront utiles que si des mécanismes de contrôle efficaces sont mis en place et qu'il existe une demande concrète d'obligation de rendre des comptes et de transparence. Il doit en effet y avoir une ferme opposition à la désensibilisation de l'ensemble des données que les consommateurs livrent en échange d'informations et de services plus pertinents. Cet échange transparent doit rester un élément central pour toute personne, ainsi que pour la société dans son ensemble.

Enfin, la triste vérité est que la rapidité de l'évolution technologique risque toujours de supplanter les efforts déployés par les législateurs quant à l'application de nouvelles technologies, comme la reconnaissance faciale, qui pose en effet de nouveaux défis. C'est la raison pour laquelle il est essentiel d'insérer des normes éthiques dans la conception de l'intelligence artificielle et des autres technologies utilisées dans l'écosystème publicitaire et au-delà. Comme l'ont fait remarquer tous les experts interrogés, la technologie est importante, mais les personnes et la culture organisationnelle le sont plus encore et elles peuvent orienter la technologie vers n'importe quel principe éthique particulier.

6.4. L'utilisation de l'intelligence artificielle pour une régulation intelligente de la publicité

Les organismes d'autorégulation de la publicité sont des organismes indépendants de normalisation de la publicité qui garantissent la conformité des communications

⁶⁴⁴ Yamakage Y. et Okamoto S., «Toward AI for human beings: Human centric AI », Zinrai. *Fujitsu scientific & technical journal.*, janvier 2017, Vol.53. no., pp. 38-44.

⁶⁴⁵ Selon une enquête de l'EDAA sur la manière dont les citoyens de l'UE perçoivent la publicité numérique depuis l'adoption du RGPD : 97 % des consommateurs savent que les données sont utilisées pour la publicité en ligne, 62 % ont une certaine compréhension de son fonctionnement et 72 % aimeraient en savoir davantage sur la manière dont les informations les concernant sont utilisées en ligne. Voir European Interactive Digital Advertising Alliance, *Consumer Research – How EU citizens perceive digital advertising since GDPR*, EDAA, <https://www.edaa.eu/consumer-research-how-eu-citizens-perceive-digital-advertising-since-gdpr/>.

commerciales au moyen de l'autorégulation et de l'application de codes publicitaires. L'Alliance européenne pour l'éthique en publicité (*European Advertising Standards Alliance* – EASA/AEEP)⁶⁴⁶ a publié un ensemble de recommandations de bonnes pratiques en matière de communications commerciales numériques, selon lesquelles toute publicité numérique diffusée en ligne relève de la compétence des organismes d'autorégulation de la publicité, c'est-à-dire que tous les formats et toutes les pratiques publicitaires numériques actuels ou en émergence, comme la promotion de produits livrés par un assistant vocal à domicile ou par l'intermédiaire de la publication d'un influenceur virtuel sponsorisé par une marque, sont soumis aux mêmes critères éthiques de conformité à la loi, de décence, d'honnêteté et d'exactitude⁶⁴⁷.

Bien que la responsabilité principale de ces organismes d'autorégulation soit de veiller à la conformité des publicités, en traitant les plaintes des consommateurs et des entreprises et en fournissant gratuitement des conseils préalables, les instances de régulation de la publicité consacrent de plus en plus leurs ressources au contrôle de la publicité assistée par ordinateur, en particulier dans l'environnement en ligne, qui connaît une augmentation annuelle à deux chiffres des dépenses publicitaires⁶⁴⁸. En fait, en 2019, deux organismes d'autorégulation de la publicité ont remporté le prix des meilleures pratiques de l'EASA pour leur application de cette technologie dans la surveillance en ligne : l'Autorité des normes publicitaires du Royaume-Uni (*Advertising Standards Authority* – ASA) a remporté un *Platinum Award* pour son projet *Avatar Monitoring*, tandis que l'organisme d'autorégulation français, à savoir l'Autorité de régulation professionnelle de la publicité (ARPP) – s'est vu décerner un *Gold Award* pour son programme d'intelligence artificielle «*Compliance as a Service*»⁶⁴⁹.

Afin d'encourager l'échange de bonnes pratiques et de contribuer à la multiplication des innovations et des connaissances en matière d'autorégulation de l'intelligence artificielle et de l'apprentissage automatique dans l'environnement réglementaire européen de la publicité, l'AESA a créé en mars 2020 un groupe de travail sur « l'autorégulation axée sur les données » (*Data Driven SR*). Les parties suivantes portent sur les projets d'autorégulation de l'intelligence artificielle en cours et examinent la manière dont ces projets, ainsi que d'autres initiatives technologiques connexes, peuvent contribuer à promouvoir une publicité digne de confiance.

⁶⁴⁶ L'AESA compte parmi ses membres 28 organismes de régulation de la publicité de ce type en Europe et 13 associations sectorielles représentant l'ensemble de la chaîne de valeur de la publicité (c'est-à-dire les marques, les agences publicitaires, les médias, les éditeurs et les plateformes numériques).

⁶⁴⁷ Alliance européenne pour l'éthique en publicité, *Best Practice Recommendation on Digital Marketing Communications*, EASA, disponible en anglais sur : <https://www.easa-alliance.org/products-services/publications/easa-best-practice-guidance>.

⁶⁴⁸ IAB Europe (2020), *AB Europe AdEx Benchmark 2019 Study*, IAB Europe, <https://iab europe.eu/knowledge-hub/iab-europe-adex-benchmark-2019-study/>.

⁶⁴⁹ Alliance européenne pour l'éthique en publicité, *Press Release: Easa's Best Practice Awards 2019*, EASA, disponible en anglais sur : <https://easa-alliance.org/news/easa/press-release-best-practice-awards-2019>.

6.4.1. Les avatars collectent des données à bon escient

En 2019, l'Autorité britannique des normes publicitaires, l'ASA, a utilisé des avatars, c'est-à-dire des logiciels de capture de données automatisés, qui reproduisaient les profils en ligne de groupes d'âge spécifiques afin de surveiller les publicités en ligne pour des produits soumis à des restrictions, tels que l'alcool, les jeux de hasard, les aliments à forte teneur en graisse, en sel ou en sucre, ainsi que les boissons gazeuses, destinés aux consommateurs britanniques.

Cette surveillance réalisée au moyen d'avatars a fait intervenir des avatars représentatifs de sept profils d'âge différents, qui ont visité 250 sites web et chaînes YouTube, et des sites connexes, et enregistré un total de 196 000 consultations de pages et saisi des informations sur plus de 95 000 publicités diffusées sur une période de deux semaines – un volume impossible à atteindre en utilisant les méthodes classiques de surveillance⁶⁵⁰. Il s'agissait là de la toute première fois que la technologie des avatars était utilisée de cette manière à des fins de surveillance ; l'ASA a ainsi recueilli un certain nombre d'informations afin de vérifier si les publicités en faveur de produits soumis à des restrictions ciblaient de manière inadéquate des publics spécifiques ou étaient diffusées dans des médias inappropriés⁶⁵¹.

Bien que la technologie automatisée employée dans ce fructueux projet pilote de surveillance automatisée n'utilisait aucun système d'intelligence artificielle, l'ASA continue à se concentrer sur l'intégration d'algorithmes d'apprentissage automatique dans les activités quotidiennes des organismes d'autorégulation de la publicité dans le cadre de leur programme à long terme *Tech4Good*, qui compte actuellement quatre projets. L'ASA a également utilisé la technologie Brandwatch, un logiciel d'intelligence sociale basé sur l'apprentissage automatique, afin de déceler les publicités illicites ou non conformes sur les médias sociaux. Lors du premier exercice de vérification, 12 000 publications problématiques qui comportaient des publicités en faveur du Botox⁶⁵² ont été supprimées sur une période de trois mois⁶⁵³.

Guy Parker, le directeur général de l'ASA⁶⁵⁴, considère que les technologies d'intelligence artificielle sont capables de renforcer l'efficacité de l'action des organismes d'autorégulation de la publicité et d'améliorer la réglementation applicable à la publicité. Il ajoute par ailleurs : « Nous estimons que les régulateurs qui n'adoptent pas l'intelligence artificielle seront à la traîne ». Il juge en effet que les applications d'intelligence artificielle

⁶⁵⁰ Alliance européenne pour l'éthique en publicité, *2018 European Trends in Advertising Complaints, Copy Advice and Pre-clearance*, EASA, disponible en anglais sur : <https://www.easa-alliance.org/products-services/publications/easa-statistics>.

⁶⁵¹ Autorité britannique des normes publicitaires (2019), *ASA monitoring report on HFSS ads appearing around children's media*, The Advertising Standards Authority Ltd, disponible en anglais sur : <https://www.asa.org.uk/resource/asa-monitoring-report-on-hfss-ads-appearing-around-children-s-media.html>.

⁶⁵² Les produits et services Botox ne peuvent pas faire l'objet de publicité auprès du public au Royaume-Uni.

⁶⁵³ Autorité britannique des normes publicitaires (2020), *Annual Report 2019*, Autorité britannique des normes publicitaires (ASA) et Commission des pratiques publicitaires (CAP), disponible en anglais sur : <https://www.asa.org.uk/news/using-technology-for-good-our-annual-report.html>.

⁶⁵⁴ Guy Parker de l'ASA. Disponible sur : <http://https://www.asa.org.uk>. Interviewé le 23 juin 2020 pour les besoins du présent article.

dans le domaine de la conformité des publicités « offrent le moyen de relever le défi de taille que pose la réglementation de la publicité en ligne », y compris pour « les nombreuses petites et moyennes entreprises qui ne sont pas toujours informées de ces dispositions et de leur importance ».

6.4.2. Les progrès de l'intelligence artificielle en matière de conformité publicitaire en France

L'Autorité française de régulation de la publicité, l'ARPP, a commencé à utiliser l'intelligence artificielle dans ses activités en 2019, dans le cadre de son programme « *Compliance as a Service* (CaaS) ». Les trois premiers projets visaient à déterminer les priorités de l'industrie concernant l'utilisation de l'intelligence artificielle au moyen d'ateliers Recherche & Développement dédiés, le lancement de l'assistant conversationnel « Jo » (un guide interactif d'assistance conversationnelle pour demander à l'ARPP des informations sur les services de pré-approbation publicitaire et des conseils en rédaction) et une application d'apprentissage approfondi pour la détection et l'analyse des boissons alcoolisées, des textes superposés et de la représentation du genre dans la publicité télévisée.

Début 2020, l'ARPP a étendu le périmètre de son dispositif d'apprentissage supervisé pour la représentation du genre - de la détection des visages à l'analyse automatique des voix et des groupes d'âge -, et a réalisé une analyse exhaustive des publicités télévisuelles soumises à l'approbation préalable de l'ARPP. Près de 131 heures de contenus vidéo ont été analysées par trois modèles d'apprentissage supervisé utilisant l'image, les technologies de reconnaissance vocale et d'identification des acteurs dans une scène, permettant à l'ARPP d'évaluer l'évolution de la représentation des hommes et des femmes⁶⁵⁵.

Mohamed Mansouri, le directeur délégué de l'ARPP⁶⁵⁶, explique que les principales raisons pour lesquelles les organismes d'autorégulation de la publicité s'intéressent aux technologies d'intelligence artificielle sont le gain de temps grâce à l'automatisation de l'apprentissage automatique des tâches à faible valeur ajoutée, ainsi que la volonté de mettre au point une autorégulation publicitaire moderne et flexible en tirant parti des dernières technologies, aussi bien pour les professionnels du secteur que pour les pouvoirs publics et les consommateurs. La consultation des représentants du secteur à l'occasion des ateliers a permis de définir plus précisément les domaines prioritaires pour les activités de surveillance de l'ARPP et pour l'évolution de son programme d'intelligence artificielle.

⁶⁵⁵ Mansouri M., *Intelligence artificielle et représentation féminine/masculine dans la publicité audiovisuelle. Après les visages et le genre : la voix et l'âge !*, blog de l'ARPP, ARPP, <https://blog.arpp.org/2020/03/05/intelligence-artificielle-et-representation-feminine-masculine-dans-publicite-audiovisuelle-visages-genre-voix-age/>.

⁶⁵⁶ Mohamed Mansouri de l'ARPP. Disponible sur : <https://www.arpp.org/>. Interviewé le 24 juin 2020 pour les besoins du présent article.

6.4.2.1. Invenio et au-delà : une étude de cas

En 2020, M. Mansouri, en collaboration avec la société Sicara⁶⁵⁷, partenaire technologique de l'ARPP, et d'autres experts des organismes français d'autorégulation de la publicité, a lancé le projet Invenio, composé de quatre modules :

- un *web-crawler* : un collecteur automatique d'annonces sur une série de sites, actuellement utilisé dans la publicité au format *display* ;
- la détection d'éventuelles infractions par l'intelligence artificielle grâce à des algorithmes basés sur l'analyse d'images et de textes ;
- le signalement des infractions et leur validation par les avocats et les experts de l'ARPP, lesquelles sont ensuite réintégrées dans le modèle afin d'améliorer sa précision, qui repose actuellement sur la détection des allégations thérapeutiques trompeuses et des publicités de produits financiers interdits ;
- un système d'alerte déclenché automatiquement lorsqu'une publicité renvoyant vers des sites interdits spécifiques est décelée.

M. Mansouri estime que le terme Invenio, qui signifie « j'ai trouvé » en latin, est parfaitement approprié pour ce projet, puisqu'il permet de gagner un temps considérable pour identifier d'éventuels manquements et s'attaquer à davantage d'infractions. Ce projet présente par ailleurs un intérêt sur le plan éducatif et permet à divers acteurs de la chaîne de valeur publicitaire de vérifier le flux de leurs programmes.

Il explique que les prochaines étapes de déploiement d'Invenio impliquent : la détection d'autres types d'infractions fréquentes dans des domaines tels que la publicité pour les produits cosmétiques et les cigarettes électroniques ; l'absence de divulgation des collaborations commerciales et autres formes de publicité concernant le marketing d'influence et les publicités vidéo. Il ajoute par ailleurs qu'Invenio englobera dans sa stratégie d'avenir la dimension européenne, « qui est plus que jamais essentielle pour démontrer la validité de l'autorégulation au niveau européen ».

Lorsqu'on lui a demandé s'il s'inquiétait du fait que les outils d'intelligence artificielle pourraient à l'avenir être trop largement utilisés et qu'ils pourraient mal évaluer et surveiller de manière excessive les publicités, M. Mansouri s'est voulu rassurant et a déclaré que « le facteur humain est fondamental dans le système » et que rien n'est décidé sans une analyse juridique préalable et l'aval des experts. « L'intelligence artificielle permet de traiter avec une grande efficacité un volume considérable de données, mais sur la base de critères très élémentaires. Même si les modèles améliorent sa précision à l'usage, le syllogisme juridique ne peut à l'heure actuelle pas être automatisé ».

⁶⁵⁷ Pour davantage d'informations sur Sicara, voir : <https://www.sicara.fr/>.

6.4.3. Tirer parti de la technologie pour renforcer la confiance du marché publicitaire néerlandais

L'organisme néerlandais de régulation de la publicité (*Stichting Reclame Code* – SRC), qui examine actuellement différentes applications possibles de technologies automatisées, est également l'un des pionniers de l'exploration de l'intelligence artificielle et de son application à l'autorégulation de la publicité. Le SRC a pour principal objectif de « renforcer la confiance dans le secteur publicitaire ». Fort de sa position dominante dans le domaine de l'autorégulation des publicités hors ligne, le SRC cherche également à se positionner dans l'environnement en ligne. Otto van der Harst, le directeur du SRC⁶⁵⁸, est convaincu que les organismes d'autorégulation de la publicité « doivent prendre des mesures et acquérir une compréhension approfondie de l'univers des technologies publicitaires » tout en conservant leur place de régulateurs indépendants. Il estime que l'organisme néerlandais de régulation et les professionnels du secteur de la publicité « se trouvent à l'aube d'une nouvelle décennie de commerce et de publicité en ligne ».

Le partenaire technologique du SRC, DEUS⁶⁵⁹, est du même avis et affirme même que « l'utilisation de la science des données et des nouvelles technologies offre une opportunité de commencer à exploiter pour de bon les énormes quantités de données de l'écosystème de la publicité en ligne ». Nathalie Post de DEUS ajoute pour sa part que parmi les principaux domaines d'action du SRC aux Pays-Bas figure la protection des personnes vulnérables (par exemple les mineurs, les personnes ayant des problèmes d'addiction et les personnes âgées) contre les contenus commerciaux potentiellement préjudiciables, comme les publicités en faveur de l'alcool, des jeux de hasard et d'argent, des régimes, de la chirurgie esthétique et du Botox, ainsi que des aliments à forte teneur en graisse, en sel ou en sucre, des boissons gazeuses, et les allégations médicales trompeuses. L'équipe de DEUS a identifié cinq domaines dans lesquels les technologies automatisées peuvent s'avérer particulièrement utiles pour épauler l'action du SRC dans ces domaines d'activité : l'utilisation d'avatars pour simuler des profils de groupes cibles spécifiques (vulnérables) ; l'utilisation d'outils d'écoute sociale pour surveiller les influenceurs des médias sociaux ; l'utilisation du traitement naturel du langage et de l'analyse d'image pour automatiser l'enregistrement, la catégorisation et le signalement des publicités qui enfreignent le code publicitaire ; l'utilisation de l'analyse d'image pour identifier les publicités qui contiennent certains préjugés, comme l'absence de diversité ; et l'acquisition d'une meilleure compréhension des flux de données dans l'écosystème des enchères en temps réel afin de lutter contre d'éventuelles infractions aux dispositions des codes publicitaires.

Otto van der Harst est persuadé que ce grand rêve d'avoir des services d'autorégulation activés par l'intelligence artificielle se profile déjà à l'horizon et que les années à venir seront consacrées à un partage d'expériences et à des discussions avec les organismes d'autorégulation de la publicité, les plateformes et les agences publicitaires, qui permettront de rallier à cette cause toutes les parties compétentes concernées :

⁶⁵⁸ Otto van der Harst du SRC. Disponible sur : <https://www.reclamecode.nl/>. Interviewé le 20 juin 2020 pour les besoins du présent article.

⁶⁵⁹ Nathalie Post de DEUS. Disponible sur : <https://deus.ai/>. Interviewée le 22 juin 2020 pour les besoins du présent article.

« L'intelligence artificielle et l'apprentissage automatique ne sont que des instruments permettant de s'assurer que l'univers en ligne reste un espace publicitaire sans danger ».

6.4.4. Les solutions technologiques des géants de l'industrie publicitaire

Les organismes de régulation de la publicité ne sont pas les seuls à traquer les publicités préjudiciables et à tenter de rendre l'espace en ligne plus sûr en utilisant des technologies d'intelligence artificielle. Certains des géants de l'industrie de la publicité en ligne, qui sont à la pointe du développement technologique, recourent également activement à des outils d'intelligence artificielle pour garantir un environnement publicitaire en ligne sûr et pérenne.

En avril 2020, Google a publié son rapport annuel sur les « publicités frauduleuses en ligne », et a affirmé avoir bloqué et supprimé 2,7 milliards de publicités illicites en 2019 (soit 400 millions de plus qu'en 2018), c'est-à-dire 10 millions de publicités par jour, 5 000 par minute et plus de 100 par seconde⁶⁶⁰.

« Nous sommes particulièrement fiers d'être une référence pour les personnes du monde entier qui recherchent des informations importantes », a déclaré l'équipe Google interrogée pour cet article⁶⁶¹. « Parallèlement à la confiance qu'ils accordent aux informations diffusées, nous savons qu'il importe que les utilisateurs puissent également avoir confiance dans les publicités qu'ils regardent ».

À mesure que l'écosystème de la publicité numérique se développe, de nouvelles menaces se profilent, nécessitant une adaptation permanente des stratégies adoptées par les entreprises et un perfectionnement de la technologie. Les experts de Google ajoutent que « Comme nous nous sommes améliorés dans la détection des tendances et des pratiques employées par les escrocs et les fraudeurs, nous ripostons en utilisant de nouvelles technologies pour faire cesser ces nouvelles menaces et en prenant davantage de mesures concernant les comptes des utilisateurs ». En 2019, ils ont redoublé d'efforts technologiques pour lutter contre les pratiques bien connues de l'hameçonnage (*phishing*) et les publicités trompeuses. Une équipe spéciale de Google a identifié et analysé les stratagèmes élaborés par les acteurs du marché peu scrupuleux pour contourner les paramètres de contrôle de Google et collecter des informations personnelles sur les utilisateurs. « Bien que l'éventail des publicités d'hameçonnage englobe les publicités qui s'adressent aux personnes intéressées par les crypto monnaies ou en quête d'informations sur le renouvellement d'un passeport en passant par les publicités qui redirigent les utilisateurs vers de faux sites bancaires en ligne, nous avons identifié un certain nombre de

⁶⁶⁰ Spencer S., *Stopping bad ads to protect users*, Google Ads, Google, <https://blog.google/products/ads/stopping-bad-ads-to-protect-users>.

⁶⁶¹ Une équipe d'experts de Google (disponible sur: <https://about.google/>) a été interrogée le 8 juillet 2020 pour les besoins du présent document, et a fourni la réponse de l'entreprise sur les sujets abordés dans ce chapitre.

dénominateurs communs sur la manière dont elles échappent à nos systèmes de contrôle, qui nous ont permis d'améliorer nos outils pour les contrer », déclare l'équipe de Google.

Les nouvelles technologies permettent en effet à Google de détecter d'éventuelles infractions, mais seule la combinaison de ces technologies et de ces compétences permettra une application effective de la loi. « Nous avons des milliers de personnes chez Google qui se consacrent à la lutte contre les publicités frauduleuses, les sites frauduleux et les escrocs en ligne », observe l'équipe. « Nous examinons en permanence les publicités, les sites et les comptes des utilisateurs pour nous assurer qu'ils respectent nos conditions d'utilisation. La mise en place de l'intelligence artificielle et d'autres technologies d'apprentissage automatique permet à Google de poursuivre son action de contrôle 24 heures sur 24 et 7 jours sur 7 et d'évaluer de multiples critères lorsqu'il décide de procéder au retrait d'une publicité.

Interrogée sur le fait que de plus en plus de publicités frauduleuses sont supprimées chaque année, l'équipe de Google a répondu que l'augmentation du nombre était révélatrice de plusieurs facteurs essentiels, notamment la nature dynamique et la dimension évolutive de l'écosystème de la publicité numérique, ainsi que les améliorations constantes dans l'adaptation aux différentes méthodes qu'utilisent les escrocs en ligne pour déjouer les systèmes de contrôle. Les experts de Google précisent que : « Aucun système ne sera jamais parfait à 100 %, mais nous sommes vigilants et nous nous efforçons en permanence de perfectionner nos outils ».

6.4.5. La future limite de l'autorégulation de la publicité

La publicité intelligente a été précédemment évoquée comme un objectif potentiel de ceux qui conçoivent et utilisent des applications d'intelligence artificielle à des fins publicitaires. Il convient par conséquent de s'interroger sur l'avenir de la réglementation de la publicité assistée par l'intelligence artificielle et sur la manière de la rendre intelligente, c'est-à-dire non seulement réactive et punitive, mais également capable d'anticiper les nouveaux enjeux et de soutenir les acteurs du secteur de manière proactive.

Les organismes d'autorégulation de la publicité interrogés s'accordent tous sur le fait que les instances de normalisation de la publicité doivent répondre aux innovations publicitaires avec leurs propres innovations technologiques. Guy Parker, de l'ASA, estime que l'intelligence artificielle et d'autres formes de technologies automatisées offriront aux organismes d'autorégulation de la publicité de meilleures indications sur la localisation et les modalités de leurs interventions ou de leurs prestations de services. Il ajoute que même si l'ASA « n'atteindra jamais la perfection », les nouvelles technologies seront à l'avenir un élément crucial de la lutte contre les publicités en ligne déloyales et préjudiciables. Mohamed Mansouri de l'ARPP affirme par ailleurs que l'intelligence artificielle permettra surtout d'aider les organismes d'autorégulation de la publicité à accroître leur efficacité, à aller toujours plus loin et à agir promptement sur le plan de la publicité numérique, où les canaux, les formats et les contenus se multiplient de manière exponentielle. Il ajoute que la collaboration dans la conception et l'utilisation de technologies d'intelligence artificielle est particulièrement importante car elle contribue au développement de l'intelligence

collective et répond aux défis inhérents à l'environnement numérique transfrontière. Otto van der Harst du SRC est du même avis et estime que les organismes de régulation de la publicité doivent maîtriser l'écosystème publicitaire s'ils « veulent être pertinents à l'avenir » et recourir aux technologies assistées pour disposer d'une forme de surveillance indépendante et être en mesure d'apporter une réponse crédible aux évolutions actuelles du marché.

Enfin, la coopération européenne jouera certainement un rôle déterminant dans la transposition des initiatives existantes, le partage d'expertise et l'apprentissage mutuel des différents acteurs du secteur. Le réseau européen des organismes de régulation de la publicité, coordonné par l'AESA, a constaté que de nouveaux formats de communications commerciales avaient perturbé le secteur de la publicité à de nombreuses reprises dans le passé, et il a fait preuve d'habileté et de souplesse pour s'adapter aux évolutions de la technologie et des sociétés, tout en respectant la diversité culturelle européenne. Grâce au partage des meilleures pratiques et à la réflexion collective sur les défis à venir, les organismes de régulation de la publicité ont continué à encourager et à soutenir l'industrie de la publicité dans la conception de publicités responsables et la préservation de la confiance des consommateurs. L'actuelle coopération sur l'utilisation de l'intelligence artificielle et d'autres technologies automatisées est l'une des initiatives les plus stimulantes à l'échelle européenne, qui fera probablement office d'incubateur pour un avenir prometteur de l'autorégulation de la publicité intelligente.

6.5. Conclusion : « La grande ruée vers les données »

De nombreux spécialistes et chercheurs qui se sont exprimés sur l'intelligence artificielle décrivent la manière dont ses applications se multiplient et s'étendent aux différents secteurs de l'industrie, parmi lesquels la publicité, et les comparent à une vague de données massives, en évoquant une puissance semblable à celle d'un tsunami qui transforme et remodèle avec force notre environnement, au point de le rendre parfois méconnaissable. Toutefois, contrairement aux tsunamis, l'opérationnalisation des données n'est pas un phénomène naturel : elle doit être organisée.

C'est la raison pour laquelle je pense qu'il serait plus pertinent d'établir une analogie entre, d'une part, les applications de l'intelligence artificielle dans le marketing et la publicité et, d'autre part, la fièvre de l'or du XIX^e siècle, qui a provoqué de profonds bouleversements dans l'économie mondiale et les modèles commerciaux, et a entraîné un mouvement migratoire et une rapide mobilité sociale. À l'image des nouveaux colons qui traversaient autrefois les océans en quête du rêve américain, l'industrie de la publicité connaît aujourd'hui une grande ruée vers les données, qui promet à ceux qui les exploitent et les canalisent efficacement d'en retirer un profit plus important encore, transformant ainsi les consommateurs en clients particulièrement lucratifs.

Tout comme les innovations mises en place à l'époque de la ruée vers l'or, les applications de l'intelligence artificielle dans la publicité ont déjà produit de nombreux avantages, aussi bien pour les professionnels du secteur que pour les consommateurs. L'intelligence artificielle a rendu la publicité plus simple et moins coûteuse quant à sa

diffusion auprès des consommateurs ciblés. Les technologies d'apprentissage automatique ont également permis d'intégrer de manière transparente de nouveaux formats publicitaires dans notre vie quotidienne, depuis les assistants vocaux qui nous proposent des produits et services sur mesure en fonction de l'historique de nos achats et d'autres facteurs déterminants, jusqu'aux influenceurs virtuels qui interagissent de manière naturelle avec nous sur les plateformes numériques, en partageant des conseils de vacances avec leurs abonnés et en encourageant l'adoption de styles de vie attrayants.

La fièvre de l'or était par ailleurs fréquemment associée au désir de s'enrichir de personnages sans foi ni loi, que symbolise le terme « Far West ». Il évoque les nouvelles techniques et les nouveaux formats de la publicité assistée par l'intelligence artificielle, qui bousculent les pratiques habituelles de ce secteur. Bien que l'on ne puisse comparer l'actuelle compétition, à laquelle se livrent les acteurs du marché pour concevoir et mettre en place les outils d'intelligence artificielle les plus efficaces qui soient, aux duels spectaculaires décrits dans les westerns, l'absence éventuelle de contrôle des technologies d'intelligence artificielle reste un défi à relever et une source de préoccupations.

L'augmentation constante de la collecte de données à caractère personnel, de l'automatisation des décisions et de l'asymétrie du marché suscitent l'inquiétude de la société civile et des décideurs politiques. L'obligation de rendre des comptes et la transparence sont primordiales si l'on veut éviter de faire naître chez les citoyens un sentiment d'aliénation à l'égard de l'intelligence artificielle et leur permettre au contraire d'adopter et de profiter pleinement des technologies qui leur offrent la possibilité de trouver plus aisément des informations pertinentes et d'obtenir plus rapidement des réponses à leurs questions, voire de les aider dans leurs projets créatifs.

En conclusion, l'intelligence artificielle risque-t-elle d'aboutir à la création d'un univers publicitaire sans foi ni loi digne de Deadwood, qui ne respecte aucune règle, qui exploite les consommateurs, et profite uniquement à ceux qui développent l'expertise technologique et s'emparent du précieux gisement de données ? Il n'en est rien, et la réalité d'un avenir aussi sombre reste très peu probable, comme d'ailleurs celle d'un univers digne du Far West. Comme nous l'avons souligné dans le présent article, la pression constante qu'exerce la société civile, ainsi que les exigences particulières des cadres réglementaires et des initiatives d'autorégulation, placent l'utilisation éthique de l'intelligence artificielle au cœur de l'actuel débat public. Les récentes évolutions politiques et les avis des experts interrogés mettent en lumière un profond désir de faire le choix d'une intelligence artificielle au service de l'humain et de son application dans la publicité. Toutefois, les stratégies et les lignes directrices doivent constamment être mises en œuvre de manière concrète et l'efficacité des mesures de protection dépend totalement de la volonté de les activer et de les appliquer ; les consommateurs ont donc une responsabilité à assumer : connaître leurs droits, agir et faire preuve de vigilance en exigeant une protection adéquate et la soumission des acteurs du marché à l'obligation de rendre des comptes.

Enfin, et peut-être surtout, la création d'un cadre durable de l'intelligence artificielle en matière de publicité et l'utilisation des données et des nouvelles technologies à bon escient sont dans l'intérêt de l'industrie publicitaire elle-même. Ce n'est qu'en intégrant des principes éthiques dans les algorithmes d'apprentissage automatique, en collectant et en exploitant les données à caractère personnel d'une manière respectueuse, qui tienne

compte de normes élevées en matière de transparence et de responsabilité, et en encourageant les applications commerciales de l'intelligence artificielle qui permettent d'identifier les acteurs peu scrupuleux du secteur et de les amener à rendre des comptes, que le marché publicitaire pourra espérer gagner durablement la confiance des consommateurs. En effet, la confiance est et restera toujours un élément incontournable en matière de publicité.

6.6. Remerciements

Tout comme les algorithmes d'intelligence artificielle, qui nécessitent un volume considérable de données pour pouvoir être exploités, le présent article n'aurait pas pu être rédigé sans les observations des experts interrogés qui ont aimablement accepté de partager leur expertise, leurs connaissances et leurs prévisions futures au sujet de l'utilisation de l'intelligence artificielle dans la publicité. Je tiens à remercier Kerry Richardson de Tiny Giants, Charlie Cadbury de Say It Now et Nathalie Post de DEUS pour avoir partagé leur expertise sur les interactions entre l'intelligence artificielle et la créativité, sur les nouveaux formats publicitaires utilisant les nouvelles technologies et sur le développement d'une approche de l'intelligence artificielle au service de l'humain. Un grand merci aux consultants indépendants en marketing d'influence Scott Guthrie et Rupa Shah de Hashtag Ad Consulting pour avoir partagé leurs observations sur le développement des influenceurs virtuels, l'éthique et l'avenir probable des humains virtuels comme canaux de marketing. Je tiens également à remercier Geraldine Proust de la FEDMA qui a eu la gentillesse de me guider à travers les cadres juridiques applicables à l'utilisation de l'intelligence artificielle dans la publicité et de me faire part de ses points de vue sur les principaux défis que poseront à l'avenir les applications d'apprentissage automatique dans le secteur de la publicité. Un immense merci à l'équipe d'experts de Google qui m'a apporté des conseils et des éclaircissements supplémentaires sur les efforts déployés par l'entreprise pour utiliser les technologies de pointe et l'expertise humaine afin de supprimer les publicités frauduleuses et protéger ses plateformes. Enfin, et surtout, je suis particulièrement reconnaissante à l'égard de mes collègues des organismes d'autorégulation de la publicité – Guy Parker de l'ASA, Otto van der Harst du SRC et Mohamed Mansouri de l'ARPP – d'avoir pris des mesures stimulantes pour adopter l'intelligence artificielle et d'autres technologies d'apprentissage automatique, afin de continuer à garantir des normes éthiques élevées dans la publicité, et de m'avoir permis de les accompagner dans ce voyage passionnant, d'apprendre en cours de route et de discuter avec eux des évolutions à venir en recherchant des solutions innovantes et en mettant l'accent sur une solide collaboration européenne.

L'intelligence artificielle dans la publicité est en quelque sorte un labyrinthe dont il est impossible de visiter tous les recoins. Toutefois, grâce aux conseils de l'ensemble des experts interrogés, j'espère avoir trouvé plusieurs parcours possibles pour y parvenir. Les parcours analysés dans ce chapitre illustrent ma propre perception de l'intelligence artificielle dans la publicité et ne sauraient être attribués à une organisation ou à une personne en particulier mentionnée dans le texte. Enfin, il convient que l'Observatoire européen de l'audiovisuel ne soit tenu responsable d'aucune des citations et opinions

présentées dans ce document, y compris celles de l'ensemble des experts interrogés et de l'auteure elle-même.

6.7. Liste des personnes interviewées

- Charlie Cadbury de Say It Now, <https://www.sayitnow.ai/>. Interviewé le 19 juin 2020.
- L'équipe d'experts de Google, <https://about.google/>. Interviewée le 8 juillet 2020 pour les besoins du présent article et qui m'a communiqué la réponse de l'entreprise.
- Scott Guthrie, consultant indépendant en gestion de marketing d'influence, conférencier et blogueur, <https://sabguthrie.info/>. Interviewé le 22 juin 2020.
- Otto van der Harst du SRC, <https://www.reclamecode.nl/>. Interviewé le 20 juin 2020.
- Mohamed Mansouri de l'ARPP, <https://www.arpp.org/>. Interviewé le 24 juin 2020.
- Guy Parker de l'ASA, <https://www.asa.org.uk>. Interviewé le 23 juin 2020.
- Geraldine Proust de la FEDMA, <https://www.fedma.org/>. Interviewée le 7 juillet 2020.
- Nathalie Post de DEUS, <https://deus.ai/>. Interviewée le 22 juin 2020.
- Kerry Richardson de Tiny Giant, <https://www.tinygiant.io/>. Interviewé le 22 juin 2020.
- Rupa Shah de Hashtag Ad Consulting, <https://www.hashtagad.co.uk/>. Interviewé le 22 juin 2020.

Droits de la personnalité

*L'IA peut être réellement créative, et son apport peut aller bien au-delà d'une simple contribution à l'élaboration d'un scénario. Après tout, le scénario ne constitue que la première étape du processus créatif. L'histoire et les idées du scénario doivent être mises en image. Dans la plupart des cas, ces histoires parlent de gens. Des gens incarnés par des acteurs. L'IA n'est pas seulement capable d'écrire un scénario et de jouer une musique, elle peut également fournir les acteurs, ou du moins transformer n'importe quel acteur en l'acteur idéal pour votre film. Elle peut, par exemple, le rajeunir ou le vieillir. Des films très récents en témoignent : dans « Gemini Man », le personnage incarné par Will Smith doit se battre contre un clone plus jeune. Une procédure de rajeunissement similaire a également été appliquée aux personnages principaux du film de Martin Scorsese, « The Irishman ». Dans Rogue One : A Star Wars Story, Carrie Fisher semble plus jeune que jamais, et Peter Cushing, décédé en 1994, y connaît également un moment de gloire post mortem. Mais généralement, après l'engouement vient l'hystérie. Si vous lisez les journaux du moment, vous pouvez tomber sur des gros titres tels que : « À l'ère des deepfakes, les acteurs virtuels peuvent-ils mettre les humains au chômage⁶⁶² » ? Imaginez en effet un film de gangsters avec comme acteur un Marlon Brando numérique, délesté de son comportement notoirement difficile sur le plateau. Quel metteur en scène refuserait cela ? Il vous faut juste, pour ce faire, disposer des bons matériels et logiciels... et d'un acteur fantôme, c'est-à-dire d'un acteur dont le visage est remplacé par celui de l'acteur plus connu : moins cher, et, dans le cas de Brando, probablement plus facile à diriger. En outre, l'IA simplifie de manière substantielle la création de figurants numériques. Comme vous pouvez l'imaginer, ces innovations tant technologiques qu'artistiques soulèvent des questions de droits de la personnalité. Ces questions juridiques sont régies par la loi et réglées par contrat. Mais un contrat peut être injuste pour la partie qui dispose d'un pouvoir de négociation plus faible, par exemple un acteur inconnu ou une personne décédée. Par ailleurs, cette pratique a également une facette plus sombre : les deepfakes. Ceux-ci peuvent être utilisés de diverses manières préjudiciables. Tout d'abord, l'exploitation commerciale illicite : les deepfakes peuvent servir à de faux parrainages de produits. Et il y a un autre problème : l'abus d'identité (le plus souvent dans des films pornographiques). Dans sa contribution à cette publication, **Kelsey Farish** souligne qu'« en particulier en ce qui concerne les technologies novatrices telles que les deepfakes et le cinéma fantôme, [...] les droits de la personnalité nécessitent une prise en compte minutieuse du contexte. »*

⁶⁶² Kemp L., « In the age of deepfakes, could virtual actors put humans out of business » », *The Guardian*, <https://www.theguardian.com/film/2019/jul/03/in-the-age-of-deepfakes-could-virtual-actors-put-humans-out-of-business>.

7. Droits de la personnalité : d'Hollywood aux *deepfakes*

Kelsey Farish, DAC Beachcroft LLP

« Si le corps n'est pas une chose, il est une situation : c'est notre prise sur le monde ».

Simone de Beauvoir

7.1. Introduction

Les juristes se tournent rarement vers les acteurs et actrices d'Hollywood pour obtenir un avis professionnel, mais lorsqu'il s'agit de technologies susceptibles de nuire à l'image publique d'un individu, l'éclairage apporté par Scarlett Johansson représente une exception notable. Lors d'une interview avec le *Washington Post*, Johansson, l'une des stars de cinéma les plus reconnaissables et les mieux payées du monde a expliqué : « il n'y a rien qui puisse empêcher quelqu'un de copier-coller mon image ou celle de n'importe qui sur un autre corps et d'obtenir un effet aussi réaliste qu'angoissant⁶⁶³ ». Bien que les effets spéciaux générés par ordinateur ne datent pas d'hier, Johansson faisait alors référence à une pratique assez peu connue à l'époque, consistant à surimposer des visages de célébrités à ceux d'acteurs de vidéos pornographiques, grâce à un système sophistiqué d'intelligence artificielle. Aujourd'hui, ce type de vidéos où des visages sont échangés, qu'elles soient ou non de nature pornographique, sont connues sous le nom de *deepfakes*.

Interrogée alors qu'elle se lançait dans une procédure judiciaire, Johansson a admis qu'il s'agissait probablement d'une « entreprise vaine, du point de vue légal » dans la mesure où « chaque pays a son propre jargon juridique sur le droit à disposer de son image ». Elle ajoutait : « et donc même s'il est peut-être possible d'obtenir le retrait de sites qui utilisent votre image aux États-Unis, l'Allemagne n'a peut-être pas de dispositions

⁶⁶³ Harwell, D., « Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image' », *The Washington Post*, www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/.

similaires⁶⁶⁴ ». Ces règles, qui visent à protéger la manière dont l'image d'un individu peut apparaître dans des publications ou des vidéos, sont souvent appelées collectivement les droits de la personnalité. Cependant, ceux-ci se composent en fait de nombreuses lois différentes qui comportent des exceptions et nuances complexes. Il est difficile d'en tirer des généralisations et souvent de prévoir l'issue des procédures judiciaires et, comme le souligne à raison Johansson, ces lois diffèrent considérablement selon les pays.

Ce chapitre a pour ambition de dissiper quelque peu la confusion qui entoure les droits de la personnalité, et de proposer un commentaire pratique sur les risques et les bénéfices dont les *deepfakes* et les acteurs fantômes s'accompagnent pour les industries du film, de la télévision, de la radiodiffusion et des médias de divertissement. La section « L'IA plante le décor » présente les technologies en question, et la section suivante sur les « droits de la personnalité et leurs implications » couvre le cadre juridique sous quatre angles différents, exemples concrets à l'appui. La section « Législation dans différents pays » revient sur le cadre juridique et décrit de quelle reconnaissance les droits de la personnalité bénéficient en Allemagne, en France, en Suède, à Guernesey, au Royaume-Uni⁶⁶⁵ et en Californie. Enfin, la dernière section « Quel avenir pour le secteur audiovisuel européen ? » s'interroge sur certaines carences importantes de ces lois ainsi que sur d'éventuelles tendances.

7.2. L'IA plante le décor : *deepfakes* et acteurs fantômes

Selon les chercheurs, l'incroyable diversité des visages humains est le fruit d'un processus évolutif visant à ce que nous soyons tous facilement reconnaissables les uns des autres, en raison de la nature très visuelle de nos interactions personnelles⁶⁶⁶. Parce que nous sommes particulièrement doués pour distinguer des visages les uns des autres, nous sommes également capables de rapidement nous apercevoir si des visages semblent étranges ou artificiels⁶⁶⁷. C'est pourquoi les scientifiques et spécialistes des effets spéciaux visuels (VFX) ont toujours eu du mal à recréer de manière fidèle l'animation des expressions du visage. Cela a changé en 2014 avec les avancées majeures d'un champ de l'intelligence artificielle connu sous le nom d'apprentissage automatique profond⁶⁶⁸.

⁶⁶⁴ Harwell D., *op.cit.*

⁶⁶⁵ Le droit du Royaume-Uni est évoqué tout au long de ce chapitre, mais le pays dispose de systèmes juridiques différents pour l'Écosse, l'Angleterre et le pays de Galles, et l'Irlande du nord. Sauf mention contraire, les références à des lois du Royaume-Uni font référence à celles qu'appliquent les tribunaux d'Angleterre et du pays de Galles, c'est-à-dire le « droit anglais ».

⁶⁶⁶ Sanders R., « Human faces are so variable because we evolved to look unique », *Berkeley News*, Université de Californie à Berkeley, <https://news.berkeley.edu/2014/09/16/human-faces-are-so-variable-because-we-evolved-to-look-unique/>.

⁶⁶⁷ Mori M., « The Uncanny Valley : The Original Essay by Masahiro Mori », *IEEE Spectrum*, traduit par K. F. MacDorman et N. Kageki [Japanese orig. 不気味の谷 (Bukimi No Tani) 1970]. Traduction anglaise disponible sur : orig. 不気味の谷 (Bukimi No Tani) 1970]. English translation available at: www.spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley.

⁶⁶⁸ Han J., « Ian Goodfellow : Invented a Way for Neural Networks to Get Better by Working Together », *MIT Technology Review*, www.technologyreview.com/lists/innovators-under-35/2017/inventor/ian-goodfellow.

7.2.1. Deepfakes

Les *deepfakes* ont commencé à gagner en notoriété lorsque des utilisateurs du site web Reddit ont partagé des vidéos pornographiques hyperréalistes de Scarlett Johansson, Gal Gadot et Emma Watson⁶⁶⁹, entre autres. Aujourd'hui, les *deepfakes* prospèrent au-delà de l'univers de la mise en image d'abus sexuels, et sont à la portée de tous en tant que forme de divertissement. Le logiciel de création de *deepfakes* est disponible en téléchargement gratuit sur des sites de partages de fichiers tels que BitTorrent et GitHub, des tutoriels sur YouTube expliquent son fonctionnement étape par étape et des créateurs freelance proposent même des *deepfakes* sur mesure pour à peine 5 euros par vidéo sur des sites de prestation de services tels que Fivver. Des applications mobiles telles que ZAO, Doublicat et AvengeThem peuvent générer des vidéos dans lesquelles les visages ont été échangés en se basant sur un seul selfie⁶⁷⁰, et même les applications grand public tels que Snapchat et Instagram ont des « filtres » qui peuvent facilement être utilisés à cet effet. La technologie continuera sans nul doute de progresser, et pour l'instant, la plupart des *deepfakes* qui ont été réalisés de façon rapide ou désinvolte sont souvent facilement identifiables lorsqu'on y regarde de plus près. Ils restent néanmoins un phénomène populaire parce qu'ils n'exigent aucune connaissance technique. De plus, leurs incohérences ou défauts mineurs ne dissuadent pas vraiment ceux qui les font seulement par curiosité ou pour s'amuser.

7.2.2. Acteurs fantômes

La modification de l'apparence de quelqu'un, par exemple pour le rajeunir ou le vieillir à des fins artistiques, est un phénomène aussi vieux que l'art d'interpréter des histoires, et les ordinateurs sont utilisés depuis les années 1970 pour animer le visage humain⁶⁷¹. Mais il a fallu attendre les années 2000 pour que des altérations du visage dignes du grand écran soient rendues possibles, comme l'ont illustré notamment *Le Seigneur des anneaux* (2001) et *Matrix Reloaded* (2003). De nombreuses techniques de VFX incluent la modélisation et le scannage du corps et du visage d'un acteur, de façon à générer un modèle virtuel connu sous le nom de double numérique. Le double numérique est alors modifié et surimposé au corps d'une doublure⁶⁷². Cette doublure peut être la même personne que celle dont le visage a été scanné au départ, et souvent, le résultat final consistera en une version plus jeune ou plus âgée de l'acteur lui-même. De façon plus controversée, le double numérique peut aussi être celui d'un acteur dont le décès a empêché une participation physique à la production.

⁶⁶⁹ Cole S., « AI-Assisted Fake Porn Is Here and We're All Fucked », *Motherboard Tech by VICE*, www.vice.com/en_us/article/qydydm/gal-gadot-fake-ai-porn.

⁶⁷⁰ Équipe Beebom, « 8 Best Deepfake Apps and Websites You Can Try for Fun », *Beebom*, www.beebom.com/best-deepfake-apps-websites.

⁶⁷¹ Parke F. I., « Computer generated animation of faces », *Association for Computing Machinery '72 Proceedings of the ACM Annual Conference*, 1: 451–457.

⁶⁷² Cosker D., Eisert P., et Helzle V., *Facial Capture and Animation in Visual Effects*, pp. 311-321 in *Digital Representations of the Real World : How to Capture, Model, and Render Visual Reality*, Université de Bath, département des sciences informatiques, http://cs.bath.ac.uk/~dpc/papers/VFX_2015.pdf.

Ces pratiques, connues sous le nom de cinéma fantôme ou hologramme, peuvent être essentielles pour assurer la continuité narrative et un rendu cohérent de la production. Au lieu de s'appuyer sur un modelage du visage obtenu lors de nouvelles séquences de jeu, les experts en VFX utilisent des prises de vue d'un acteur décédé pour créer son double numérique, qui est à son tour modifié et surimposé au corps d'une doublure. À titre d'exemple récent, à la suite du décès prématuré de l'actrice Carrie Fischer en 2016, Lucasfilms a utilisé des prises de vues d'archives de l'actrice pour qu'elle puisse à nouveau incarner la princesse Leia dans le film de 2019, *Star Wars: L'Ascension de Skywalker*. Notons tout de même que l'un des responsables du studio avait alors assuré n'avoir « aucune intention de lancer une mode visant à recréer des acteurs disparus⁶⁷³ ».

7.3. Les droits de la personnalité et leurs implications

Les droits de la personnalité ont pour objectif de donner aux individus la possibilité de contrôler la publication de leur image, ou de certaines de ses caractéristiques. Les caractéristiques protégées incluent généralement des photographies et des images du visage et du corps, mais elles peuvent aussi comprendre des noms et des styles, signatures, voix, ou même une gestuelle qui seraient distinctifs. Les accessoires fortement associés à l'individu peuvent également faire l'objet d'une protection, comme l'a illustré en Italie une affaire dans laquelle l'usage à des fins publicitaires du chapeau et des lunettes d'un chanteur populaire, sans son consentement, a été considéré comme une atteinte à son image⁶⁷⁴. Pris dans leur ensemble, ces attributs constituent le « personnage » ou l'« image » d'un individu, l'image désignant ici l'image publique plutôt qu'une photographie. C'est pour cette raison que certains systèmes juridiques, comme ceux au Royaume-Uni ou en France, désignent les droits de la personnalité sous le nom de « droits à l'image ». Mais quel que soit leur dénomination, ils peuvent poser un certain nombre de curieuses complications juridiques pour le secteur audiovisuel.

Dans la mesure où la technologie et la société ont évolué plus rapidement que le droit, les juristes sont contraints de se tourner vers des doctrines éprouvées et plus anciennes pour régler les litiges relatifs au droit à l'image d'un individu. Il est souvent dit des droits de la personnalité qu'ils ont deux facettes : le droit à l'image et le droit au respect de la vie privée. Le droit à l'image couvre l'utilisation de l'image d'un individu pour des motifs commerciaux, et le droit au respect de la vie privée cherche à empêcher une intrusion injustifiée dans la vie intime et familiale d'un individu.

Il convient de s'arrêter un instant pour rappeler que le droit est un langage nuancé dont les règles obscures et la jurisprudence ancienne ont une nature quelque peu mythologique qui reflète les normes sociales à partir desquelles il a été élaboré. Le bon usage des droits de la personnalité dépend du contexte ainsi que de la culture, et il n'existe pas d'approche commune, même au sein de l'Union européenne. D'une certaine façon, ce que l'on observe du point de vue juridique est comparable à une boîte de nuit un peu

⁶⁷³ Robinson J., « Star Wars : The Last Jedi – What Happened to Leia ? », *Vanity Fair*, www.vanityfair.com/hollywood/2017/12/star-wars-the-last-jedi-does-leia-die-carrie-fisher-in-episode-ix.

⁶⁷⁴ *Dalla c. Autovox SpA Pret.* Di Roma, 18 avril 1984, Foro It. 1984, I, 2030, Giur. It. 1985, I, 2, 453.

sombre, dans laquelle les lois portant sur la propriété intellectuelle, le marketing et la publicité, la gestion de la réputation, la liberté d'expression, la détresse émotionnelle et la protection du consommateur danseraient ensemble sur la même piste, mais pas forcément en harmonie. Le reste de cette section cherche à mettre un peu d'ordre dans cette danse, en abordant les droits de la personnalité sous quatre angles différents.

7.3.1. Angle 1 : L'image comme propriété (intellectuelle)

Les stars de cinéma, les athlètes professionnels et autres célébrités peuvent gagner des sommes d'argent importantes au-delà de leur travail quotidien au studio ou au stade. Les usages lucratifs de leur *personnalité* peuvent inclure des publicités rémunérées, des marchandises officielles ou des collaborations avec des maisons de couture. Ce n'est pas seulement l'individu dont les intérêts financiers sont en jeu : son équipe de gestion, ses sponsors, les studios de cinéma, les labels de musique et bien d'autres ont également un intérêt à la valeur monétaire de sa *personnalité*. Dans ce contexte, ce premier angle considère la question du droit à l'image des droits de la personnalité comme une forme de protection économique.

Selon la théorie de la propriété fondée sur le travail, un individu a le droit de jouir des fruits de son labeur⁶⁷⁵. La propriété intellectuelle est donc, par extension, un droit qui découle du travail intellectuel et créatif. Les droits d'auteur sont une forme de propriété intellectuelle qui reconnaît l'effort et l'investissement qu'implique la production d'œuvres de création, qu'il s'agisse de longs métrages ou de clips sur Tiktok, ou de tout ce qui se trouve entre ces deux extrêmes. Les lois en matière de droit d'auteur interdisent à d'autres de copier, de modifier ou d'utiliser ces créations sans la permission de leur créateur et, dans l'écosystème en ligne, les plaintes pour atteinte au droit d'auteur sont devenues un moyen répandu, pour les créateurs et les studios, d'affirmer un contrôle sur leurs contenus.

Malheureusement, cette approche des droits de la personnalité sous l'angle des atteintes à la propriété intellectuelle ne va pas sans poser plusieurs défis pratiques. En premier lieu, une complication quelque peu paradoxale peut survenir lorsqu'il s'agit de faire valoir des protections du droit d'auteur au nom de la personne représentée. En effet, il faut se rappeler que c'est le travail du photographe ou du vidéaste qui est protégé par le droit d'auteur, et ces droits sont souvent directement cédés par le créateur au studio ou à la société de production du film en question. C'est pourquoi il est impossible pour un individu de faire valoir une plainte en matière de droit d'auteur simplement du fait qu'il apparaît dans un contenu audiovisuel. Même si certains détenteurs de droits pourraient accepter de lancer une procédure judiciaire pour le compte d'une personne représentée, cela ne sera pas toujours le cas, en particulier si le différend oppose le détenteur de droit et l'interprète lui-même.

Des centaines voire des milliers d'images ont pu être utilisées pour générer des doubles numériques. Une fois ces images fusionnées, l'identification des détenteurs des

⁶⁷⁵ Hughes J., « The Philosophy of Intellectual Property », *Centre du droit de l'Université de Georgetown et Georgetown Law Journal*, 77 Geo. L.J. 287.

droits relatifs aux images originales peut être difficile, voire impossible. De plus, certains usages de matériel protégé par le droit d'auteur sont licites en dépit de l'absence d'autorisation. Ces exceptions incluent notamment la satire et la parodie, tandis qu'aux États-Unis, la transformation d'une œuvre créative pour générer une signification ou des expressions nouvelles peut également être acceptable.

7.3.2. Angle 2 : Image et reconnaissance de la marque

À l'instar du premier angle, ce deuxième angle porte sur l'utilisation commerciale de l'image d'une personne, mais sans considérer l'image d'un individu dans un contenu audiovisuel du point de vue de la propriété. Au lieu de cela, selon le principe de l'enrichissement sans cause (ou enrichissement illégitime ou injustifié), il est interdit de tirer injustement un avantage financier de la notoriété ou de la réputation d'autrui. De la même manière que des symboles et des noms peuvent être déposés en tant que marques pour différencier ses marchandises et services de ceux de la concurrence, il est possible de protéger l'image d'une personne quand cette image est utilisée pour indiquer une certaine qualité, provenance ou affiliation. Mais, à la différence de l'enregistrement d'un symbole ou d'une phrase autonome, il est impossible de déposer à titre de marque tous les aspects du *personnage* d'une célébrité. Tout d'abord, une séquence vidéo ne peut pas faire l'objet d'une protection en tant que marque. Deuxièmement, même si un plan particulier ou une image spécifique a été déposé en tant que marque, cela n'empêchera pas qui que ce soit d'utiliser des photographies différentes pour créer un *deepfake* ou une scène de cinéma fantôme. Afin de combler ce vide juridique, certains pays cherchent à réparer le préjudice financier découlant de l'usage abusif d'une image.

Cela est particulièrement vrai s'agissant d'affaires de parrainages trompeurs ou faux, comme par exemple lorsque l'enseigne de mode britannique Topshop a mis en vente un tee-shirt à l'effigie de la chanteuse Rihanna, titulaire de plusieurs Grammy awards. Topshop disposait bien d'une licence de droit d'auteur pour l'usage de cette image, mais Rihanna a démontré que le tee-shirt portait préjudice à l'attractivité de sa marque personnelle et elle a donc obtenu gain de cause contre Topshop pour utilisation abusive de son image⁶⁷⁶. Au cours d'un autre procès qui portait là aussi sur l'usage de la photo d'une popstar sur des tee-shirts, la Cour fédérale allemande a accordé un dédommagement à la chanteuse de *99 Luftballons* Nena car elle n'avait pas été rémunérée pour l'utilisation de cette image particulière⁶⁷⁷. Il faut souligner que dans certains pays, cette reconnaissance de marque ne peut entrer en jeu que si la personne est en mesure de démontrer le préjudice financier porté à son image de marque. Cela laisse évidemment beaucoup à désirer pour les personnes moins connues ou celles qui ne peuvent démontrer la valeur monétaire de leur *personnage*.

⁶⁷⁶ *Robyn Rihanna Fenty c. Arcadia Group Brands Ltd (t/a Topshop)*, Court of Appeal (Civil Division) - [2015] EWCA Civ 38.

⁶⁷⁷ *Nena*, BGH, Urt. V. 14 octobre 1986 VI ZR 10/86 Oberlandesgerichtsbezirk Celle, Landgericht Lüneburg.

7.3.3. Angle 3 : Protections du droit à la vie privée

Les deux angles examinés jusqu'ici portaient sur la manière dont les lois en matière de droit à l'image peuvent protéger l'usage de l'image d'un individu à des fins publicitaires ou commerciales. Ce troisième angle aborde la façon dont les lois en matière de respect de la vie privée fonctionnent pour protéger contre une intrusion injustifiée dans la vie privée ou familiale d'autrui, et comment ces lois pourraient s'appliquer aux *deepfakes* et aux acteurs fantômes. Tandis que les droits de propriété portent sur la jouissance d'un actif tangible ou intellectuel, le droit au respect de la vie privée découle simplement de la qualité d'être humain.

La Convention européenne des droits de l'homme (ci-après la « Convention »)⁶⁷⁸ est l'instrument central qui protège les droits de l'homme et les libertés politiques en Europe. Elle s'applique à tous les États membres de l'Union européenne ainsi qu'à 20 autres pays. L'article 8 de la Convention confère à toute personne un droit fondamental au respect de la vie privée, qui comprend la protection contre des intrusions injustifiées dans leur vie personnelle. Comme le savent les amateurs d'émissions documentaires ou de télé réalité, les images permettent au public de devenir des spectateurs et même des voyeurs de la vie des personnes représentées⁶⁷⁹. Les photographies et séquences filmées sont donc traitées avec prudence par la justice, et font l'objet d'un examen minutieux par les tribunaux.

Malgré cela, le droit au respect de la vie privée ne signifie en aucun cas une protection absolue de son image. La plupart des sociétés considèrent qu'il convient de rechercher un équilibre entre le respect de la vie privée et d'autres droits concurrents, lesquels l'emporteront parfois. Par exemple, les médias d'information auront le droit – et parfois le devoir – de partager des photographies « privées » de personnes célèbres, lorsque cela est dans l'intérêt public. Bien que le droit au respect de la vie privée soit un droit fondamental en vertu de la Convention, cela est également vrai du droit à la liberté d'expression découlant de son article 10. La plupart des pays qui ne sont pas parties à la Convention bénéficient d'une disposition similaire : aux États-Unis, il s'agit principalement du premier amendement à la Constitution⁶⁸⁰.

Ainsi, dans la plupart des affaires portant sur le droit à la vie privée, l'analyse prendra en compte si l'image a été prise en public ou si l'individu était à ce moment-là raisonnablement en droit de s'attendre à une protection de sa vie privée. Une photo obtenue par des paparazzis au moyen d'un téléobjectif, sur laquelle figurerait une célébrité allongée et seins nus au bord de la piscine de sa villa sécurisée, sera naturellement traitée d'une façon différente d'une photo d'elle vêtue de la tête aux pieds sur le tapis rouge d'un festival de cinéma. Le problème qui se pose du point de vue du droit au respect de la vie privée est cependant qu'en pratique, les millions de photographies prises et partagées quotidiennement se situent généralement quelque part entre ces deux extrêmes. Il n'est donc pas simple de prévoir quelles images seront considérées comme une intrusion dans

⁶⁷⁸ Convention européenne des droits de l'homme (formellement Convention pour la sauvegarde des droits de l'homme et des libertés fondamentales), www.echr.coe.int/Documents/Convention_FRA.pdf.

⁶⁷⁹ *Michael Douglas, Catherine Zeta-Jones & Ors c. Hello! Ltd. & Ors* [2005] EWCA Civ 595.

⁶⁸⁰ Premier amendement à la Constitution des États-Unis, www.constitution.congress.gov/constitution/amendment-1/.

la vie privée d'une personne ou comme un usage abusif d'informations confidentielles la concernant. De plus, selon une opinion répandue, les célébrités qui tirent un avantage financier de leur notoriété devraient nécessairement renoncer dans une certaine mesure à leur droit au respect de la vie privée. Que cela soit juste ou non est une autre question.

Lorsqu'un contenu audiovisuel divulgue des informations confidentielles, par exemple si l'on y entend une personne révéler des faits du domaine privé, il pourrait s'agir d'une intrusion dans sa vie privée. Mais, au-delà de ce scénario, les lois en matière de respect de la vie privée ne constituent sans doute pas des outils appropriés pour lutter contre les *deepfakes* ou des recours indésirables à des acteurs fantômes. D'abord, de nombreuses images et vidéos sont souvent fusionnées pour créer un double numérique, et il serait sans doute vain de tenter de déterminer si l'une d'entre elles avait au départ un caractère privé. Deuxièmement, et aspect peut-être plus important encore, les *deepfakes*, par définition, mettent en scène quelque chose qui ne s'est jamais produit, et des scénarios imaginaires ne sauraient constituer une atteinte à la vie privée. Et en ce qui concerne l'utilisation de scènes tournées par un acteur désormais disparu, il est fort probable que ses héritiers ou successeurs auraient bien du mal à faire valoir un droit au respect de la vie privée post mortem. La Cour européenne des droits de l'homme a toujours hésité à reconnaître un droit à la vie privée des personnes décédées, sauf lorsque que celui-ci était connecté à celui de personnes vivantes⁶⁸¹. Dans quasiment tous les cas de figure, une personne disparue ne dispose pas d'un droit au respect de la vie privée auquel il pourrait être porté atteinte à proprement parler, et, dans la mesure où les droits à la vie privée sont des droits inaliénables et inséparables de la personne en cause, ils ne sauraient faire l'objet d'un transfert ou d'une transmission.

7.3.4. Angle 4 : La dignité et les droits voisins

Les perspectives examinées jusqu'ici ont tenté d'apporter un éclairage utile des aspects liés au droit à l'image et au droit au respect de la vie privée des droits de la personnalité. Bien que les lois mentionnées plus haut apportent des protections parfois appropriées, elles n'empêchent toutefois pas la possibilité d'abus. Une solution pourrait consister à élargir les droits de la personnalité pour y inclure le bien-être émotionnel et mental, indépendamment de toute considération financière ou relative au respect de la vie privée. Ce quatrième et dernier angle se focalise donc sur le principe selon lequel toute personne dispose d'un droit fondamental à la dignité et à l'intégrité personnelles. Il est également possible de faire valoir que les lois qui protègent actuellement l'intégrité des interprétations enregistrées pourraient également s'appliquer aux performances virtuelles.

La Convention de Genève reconnaît que le « respect de l'être humain et de sa dignité est un principe universel qui s'impose même en l'absence de tout engagement

⁶⁸¹ Buitelaar, J.C., « Post-mortem privacy and informational self-determination », *Ethics and Information Technology*, 19(2), pp.129–142, <https://link.springer.com/article/10.1007/s10676-017-9421-9>.

contractuel »⁶⁸². Cette reconnaissance explicite de la dignité est inscrite dans la constitution d'après-guerre de l'Allemagne et dans la législation et la jurisprudence de la plupart des pays européens. Il est donc raisonnable d'avancer que ces lois ont été inspirées par le souhait de protéger l'intégrité, l'individualité et l'autodétermination des personnes⁶⁸³. À titre d'exemple, l'affaire *n° 2 von Hannover c. Allemagne*⁶⁸⁴, souvent citée en matière de vie privée, concernait la Princesse Caroline de Hanovre, la fille aînée du Prince Rainier III de Monaco. La Princesse Caroline tentait depuis longtemps de se protéger contre la publication de photos d'elle par la presse allemande. Lorsque des photographies d'elle et de sa famille ont été publiées sans son accord, l'affaire a fini par atterrir devant la Cour européenne des droits de l'homme. Dans sa décision, la Cour a déclaré que « l'image d'un individu est l'un des attributs principaux de sa personnalité, du fait qu'elle exprime son originalité et lui permet de se différencier de ses pairs. Le droit de la personne à la protection de son image constitue ainsi l'une des conditions essentielles de son épanouissement personnel ».

Évidemment, internet a considérablement facilité la publication de contenus audiovisuels susceptibles d'embarrasser et d'humilier. La nature discursive des médias sociaux rend ces contenus encore plus visibles, ajoutant encore à l'embarras subi. En plus de porter préjudice à l'estime de soi et à la santé mentale de la personne représentée, de telles vidéos pourraient avoir un effet domino important sur ses partenaires à l'écran ou ses partenaires commerciaux. Bien qu'il s'agisse là d'un domaine émergent du droit, les arguments en faveur d'une solution juridique centrée sur la dignité mettent en avant une vision plus globale de l'identité. En temps utile, la reconnaissance légale de cette facette de la personnalité pourrait amener à moderniser les droits de la personnalité, et ainsi protéger les individus des préjudices découlant de l'abus de ces technologies⁶⁸⁵.

Il est néanmoins quelque peu frustrant que, lors des procès qui mettent en jeu l'honneur et la tranquillité d'un individu, la notion de dignité soit souvent simplement abordée comme un principe générique se référant à un idéal abstrait. Même lorsque la loi souligne l'importance de la dignité humaine, il n'est pas rare que les universitaires et les juristes en exercice peinent à définir des règles cohérentes pour le fonctionnement pratique de ce droit. La résolution de nombreuses affaires, y compris *von Hannover*, dépend au bout du compte des lois en matière de diffamation ou de respect de la vie privée. Comme nous l'avons vu ci-dessus, il faudra, dans ces affaires, que les droits de l'individu ayant subi un préjudice prévalent sur le droit à la liberté d'expression des éditeurs.

Il reste toutefois un mince espoir pour ceux qui travaillent dans les arts du spectacle. Tandis que le droit d'auteur cherche à protéger les intérêts de l'auteur ou du créateur d'une œuvre, les droits voisins portent sur la nécessaire contribution des musiciens, danseurs ou

⁶⁸² Préambule de la Convention de Genève, Convention (IV) relative à la protection des personnes civiles en temps de guerre. Genève, 12 août 1949,

<https://ihl-databases.icrc.org/applic/ihl/dih.nsf/Comment.xsp?action=openDocument&documentId=7B16B56E1A89C8CC12563BD002CEEB3>.

⁶⁸³ Abraham, K. et White E., « The Puzzle of the Dignitary Torts », *Cornell Law Review* 104 (2), pp.317–380, www.core.ac.uk/download/pdf/228302795.pdf.

⁶⁸⁴ *Von Hannover c. Allemagne (n° 2)* 40660/08 [2012] Cour européenne des droits de l'homme 228.

⁶⁸⁵ Dunn S., *Identity Manipulation : Responding to advances in artificial intelligence and robotics*, document de travail présenté à la conférence WeRobot d'avril 2020. Le document a été fourni à Kelsey Farish dans le cadre d'une correspondance privée en juillet 2020 et est cité ici avec l'aimable autorisation de son auteur.

acteurs lors d'une performance enregistrée. Ces droits voisins sont ainsi nommés car ils sont proches de la notion de droit d'auteur (ils sont également connus sous le nom de « droit des artistes-interprètes » pour les distinguer du « droit d'auteur »). Les droits voisins ont généralement pour but d'assurer la reconnaissance de la contribution d'un artiste, ainsi que d'empêcher une modification d'un enregistrement susceptible de porter préjudice aux intérêts intellectuels ou personnels de l'artiste-interprète. Par essence, les droits voisins traitent des problèmes découlant de la copie ou de la diffusion pirate d'émissions sans compensation appropriée des interprètes impliqués.

Malgré leur importance évidente en ce qui concerne les performances diffusées par un média audiovisuel, les droits voisins sont très rarement abordés dans le contexte des droits de la personnalité. Généralement, les litiges en matière de droit de la personnalité portent sur l'usage abusif de photographies statiques, qui peuvent facilement être altérées pour obtenir un effet réaliste, plutôt que sur des performances enregistrées. Mais comme nous l'avons vu, les avancées de la VFX et de l'intelligence artificielle font que des vidéos manipulées sont désormais davantage susceptibles d'être prises pour des vidéos authentiques. Cette ère nouvelle, où sont créées des œuvres audiovisuelles qui imitent des performances réelles, pourrait justifier une mise à jour de la définition d'une « performance » traditionnellement retenue par la législation en matière de droits voisins.

Il existe en Californie des précédents en matière de reconnaissance de performances virtuelles non consenties, dans lesquels des tribunaux ont indemnisé des individus pour leur « asservissement » numérique dans des jeux vidéo. Les musiciens du groupe No Doubt ont par exemple gagné un procès contre les créateurs de Band Hero, un jeu qui permet aux joueurs de choisir un avatar numérique hyperréaliste pour simuler une participation à la performance d'un groupe de rock⁶⁸⁶. Si les *deepfakes* ou le cinéma fantôme étaient considérés comme des « performances » dans le cadre de la loi, alors l'image d'une personne qui aurait été contrainte de se plier à une performance numérique contre son gré pourrait être protégée. C'est pourquoi, afin d'avoir une vue complète de ces questions, les droits voisins devraient être pris en compte au même titre que tous les autres droits à l'image ou au respect de la vie privée qui composent le dispositif en matière de droits de la personnalité.

7.4. Législation dans différents pays

En résumé, comme l'explique la section 2, les droits de la personnalité sont souvent répartis en deux grandes catégories. La première, considérée comme le droit à l'image, concerne l'exploitation de la notoriété et de la réputation d'un individu pour en tirer un gain financier. Le droit à l'image peut être vu sous plusieurs angles différents mais connexes, à savoir, les droits en matière de propriété intellectuelle (angle 1 ci-dessus) et les lois en matière de protection des marques (angle 2 ci-dessus). De façon distincte des questions d'image commerciale, la deuxième catégorie consiste à protéger la vie privée des individus en interdisant la publication de certaines images au moyen de lois relatives au respect de la

⁶⁸⁶ *No Doubt c. Activision Publishing, Inc.*, 122 Cal.Rptr.3d 397 (Cal. Ct. App. 2001).

vie privée (angle 3 ci-dessus). La dignité en tant qu'aspect de la personnalité a peu retenu l'attention des spécialistes des droits de la personnalité, et les droits voisins encore moins (angle 4 ci-dessus). La question de savoir si ces droits s'appliquent après le décès d'un individu donne lieu à diverses réponses, mais les droits post mortem seront généralement considérés comme une forme de propriété qu'un individu peut transmettre à ses héritiers.

Il reste à voir comment les prochains litiges sur les *deepfakes* et le « jeu fantôme » seront soldés. S'agissant du moins des pays choisis dans cette section 3, ce sont les lois relatives à la réputation et à la reconnaissance de marque qui seront vraisemblablement les plus pertinentes. Les droits à la dignité et les droits voisins pourraient également s'appliquer, et pourraient même peut-être constituer un cadre plus moderne pour la protection d'un *personnage*. Par ailleurs, les droits de propriété intellectuelle et au respect de la vie privée sont peut-être moins importants lorsqu'il s'agit d'images détournées d'individus figurant dans des contenus audiovisuels. Dans tous les cas, l'image, la vie privée et la dignité d'un individu nécessitent une analyse au cas par cas, ce qui impliquera toujours de trouver un équilibre entre ces droits et les droits concurrents en matière de liberté d'expression.

7.4.1. Allemagne

L'Allemagne abrite le plus grand marché audiovisuel d'Europe, et son secteur des médias atteignait une valeur de 23 milliards d'euros en 2018⁶⁸⁷ et employait plus de 520 000 personnes en 2017⁶⁸⁸. La reconnaissance des droits de la personnalité y est large et ancrée dans la Constitution allemande. Celle-ci fournit un exemple presque parfait des protections en matière de dignité évoquées sous l'angle 4 ci-dessus, en affirmant que « la dignité humaine est inviolable » et que « chacun a droit au libre épanouissement de sa personnalité » (Grundgesetz, articles 1 et 2)⁶⁸⁹. Ces protections fondamentales sont complétées par le Code civil, selon lequel toute personne qui porte abusivement préjudice aux droits de la personnalité d'un individu est responsable des dommages qui en résultent (BGB, §823)⁶⁹⁰. En Allemagne, la dignité d'une personne peut subsister après la fin de sa vie

⁶⁸⁷ Observatoire européen de l'audiovisuel, Annuaire 2019/2020 Tendances clés – Cinéma, télévision, vidéo et services audiovisuels à la demande – le paysage paneuropéen. Observatoire européen de l'audiovisuel, <http://yearbook.obs.coe.int/s/document/key-trends/2019>.

⁶⁸⁸ Weidenbach B., « Beschäftigte in der Medienbranche in Deutschland 2017 », *Statista*.

⁶⁸⁹ Grundgesetz, formellement Grundgesetz für die Bundesrepublik Deutschland [loi fondamentale de la République fédérale d'Allemagne]. Traduction anglaise disponible sur : www.gesetze-im-internet.de/englisch_gg/englisch_gg.html.

⁶⁹⁰ BGB, formellement Bürgerliches Gesetzbuch [Code civil allemand] dans sa version promulguée le 2 janvier 2002 (Journal officiel fédéral, page 42, 2909 ; 2003 page 738) amendé en dernier lieu par l'article 4, paragraphe 5 de la loi du 1^{er} octobre 2013. Traduction anglaise disponible sur : www.gesetze-im-internet.de/englisch_bgb/englisch_bgb.html.

naturelle, et les proches peuvent ainsi protéger une personne disparue contre un discrédit porté à son image⁶⁹¹.

Tous les aspects de l'identité d'une personne sont couverts par le droit allemand consolidé de la personnalité, mais les images et photographies d'un individu font l'objet de protections renforcées. La loi allemande relative au droit d'auteur prévoit que, lorsque des images incluent un individu identifiable, elles « ne sauraient être distribuées ou exposées publiquement qu'avec le consentement de la personne en question » (UrhG, §19a et §22)⁶⁹². Le caractère identifiable porte non seulement sur des représentations du visage, mais aussi d'autres caractéristiques distinctives, sur des dessins humoristiques ou même des sosies (*doppelgänger*). Il faut souligner que la notion de distribution est interprétée de façon large, et donc même un photographe amateur est théoriquement censé obtenir un consentement avant de poster l'image d'une personne sur les médias sociaux, ou de la partager avec des amis sur une application de messagerie. Le consentement est également requis d'un artiste-interprète avant la diffusion ou toute autre communication de la performance enregistrée (Section III, UrhG).

Le consentement n'est toutefois pas nécessaire dans le cadre de la représentation de personnes de l'histoire contemporaine, ou d'images dans lesquelles les individus ne figurent que de manière accessoire, ou lorsque la publication est au service d'un intérêt supérieur artistique ou culturel (KunstUrhG, §23)⁶⁹³. Mais cet exercice d'équilibrage a récemment évolué en faveur de la protection du détenteur des droits sur l'image⁶⁹⁴. Et en ce qui concerne les images particulièrement intimes, par exemple de quelqu'un dans son domicile, le code pénal interdit de prendre, transmettre ou utiliser des images qui portent atteinte à la vie intime d'autrui, et les violations sont passibles d'une amende ou d'un maximum de deux ans de prison (StGB, §201a).⁶⁹⁵

S'agissant de la question du droit à l'image dans le cadre des droits de la personnalité, les tribunaux allemands sont de plus en plus disposés à protéger les individus contre une exploitation commerciale indésirable de leur image⁶⁹⁶. Au contraire de l'approche constatée au Royaume-Uni, une commercialisation préalable du *personnage* d'un

⁶⁹¹ Seyfert C., *Regional Court Frankfurt am Main : Postmortem personality right of a Holocaust survivor prevails over publications by a British history professor (Affaire 2-03 O 306/19)*, cabinet d'avocats Zeller & Seyfert, www.zellerseyfert.com/en/litigationblog-detail/items/regional-court-frankfurt-am-main-postmortem-personality-right-of-a-holocaust-survivor-prevails-over-publications-by-a-british-hi.html.

⁶⁹² UrhG, formellement Urheberrechtsgesetz [loi sur les droits d'auteur et les droits voisins] du 9 septembre 1965 (Journal officiel fédéral I, p. 1273) amendée en dernier lieu par l'article 1 de la loi du 28 novembre 2018 (Journal officiel fédéral I, p. 2014). Traduction anglaise disponible sur : www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html.

⁶⁹³ KunstUrhG, formellement Gesetz betreffend das Urheberrecht an Werken der bildenden Künste und der Photographie [loi sur le droit d'auteur dans les œuvres de beaux-arts et de photographie] du 9 janvier 1907 (RGL I p. 7) amendée en dernier lieu par la loi du 16 février 2001 (BGBL I, p. 266), 1^{er} août 2001, www.gesetze-im-internet.de/kunsturhg/_23.html.

⁶⁹⁴ Coors C., « Image Rights of Celebrities vs. Public Interest – Striking the Right Balance Under German Law », *Journal of Intellectual Property Law & Practice* (2014) 9 (10): 835-840, www.ssrn.com/abstract=2738514.

⁶⁹⁵ StGB, formellement Strafgesetzbuch [Code pénal] dans sa version promulguée au 13 novembre 1998 (Journal officiel fédéral I p. 3222) amendé en dernier lieu par l'article 3 de la loi du 2 octobre 2009 (Journal officiel fédéral I p. 3214). Traduction anglaise disponible sur : www.gesetze-im-internet.de/englisch_stgb/.

⁶⁹⁶ Peters M., « The Media and Entertainment Law Review – Germany », *The Law Reviews*, www.thelawreviews.co.uk/edition/the-media-and-entertainment-law-review-edition-1/1211744/germany.

individu n'est pas une condition expresse pour pouvoir faire valoir un droit à l'image. S'agissant de la représentation d'une personne disparue, un consentement des proches est requis pendant les dix ans suivant son décès. Dans une affaire concernant des photos de la star de cinéma Marlene Dietrich (décédée plusieurs années auparavant), la Cour fédérale de justice a estimé que toute exploitation non autorisée d'images ouvrait un droit à compensation pour le détenteur du droit à l'image, et ce, quelle que soit la gravité de la violation⁶⁹⁷.

7.4.2. France

Généralement considérée comme le berceau du cinéma, la France dispose, selon de nombreux indicateurs, de l'industrie cinématographique la plus prolifique d'Europe : près de 300 films par an ou plus ont été réalisés en France chaque année depuis 2015⁶⁹⁸. Les droits de la personnalité en France, connus sous le nom de droit à l'image ou à son image, incluent des lois relatives à la protection de la vie privée qui ont pour but de protéger les individus d'une exposition non souhaitée, ainsi que des droits commerciaux qui permettent à l'image d'être exploitée en tant qu'actif commercial⁶⁹⁹. En règle générale, le consentement de la personne doit être obtenu avant que son image soit communiquée au public. Comme d'autres pays, la France définit l'image de façon large pour englober l'apparence, la voix, la photographie, le portrait ou la vidéo. Les tribunaux français ont confirmé que le floutage du visage d'un modèle ne saurait en lui-même résoudre une atteinte au droit de l'image, lorsque d'autres parties du corps sont encore visibles⁷⁰⁰.

Cette philosophie est largement ancrée dans la forte protection garantie en France de la vie privée, intime ou familiale des personnes. Le Code civil affirme que chacun a droit au respect de sa vie privée et, surtout, il confère aux tribunaux la possibilité de recourir à toute mesure appropriée pour empêcher ou faire cesser une atteinte à l'intimité de la vie privée (Code civil, article 9)⁷⁰¹. Ainsi, les juges français ont progressivement développé au cas par cas des droits de la personnalité plus protecteurs pour les plaignants⁷⁰². La nature du consentement constitue une question centrale, mais elle n'est pas toujours facile à déterminer. Le tribunal de grande instance de Paris a récemment estimé qu'un mannequin professionnel participant à une séance photo à des fins publicitaires n'avait pas consenti à toute forme d'exploitation commerciale de ces images⁷⁰³.

⁶⁹⁷ Marlene Dietrich, BGH 1^{er} décembre 1999 - 1 ZR 49/97 - Kammergericht LG Berlin.

⁶⁹⁸ Lemerrier F., « 301 feature films produced by France in 2019 », *Cineuropa - the best of European cinema*, <https://cineuropa.org/en/newsdetail/387425/>.

⁶⁹⁹ Logeais E. et Schroeder J-B., « The French Right of Image : An Ambiguous Concept Protecting the Human Persona », *18 Loyola University Entertainment Law Review* 511, <https://digitalcommons.lmu.edu/cgi/viewcontent.cgi?article=1366&context=elr&httpsredir=1&referer=>.

⁷⁰⁰ *M. X c. Umanlife*, TGI de Paris, décision du 16 novembre 2018.

⁷⁰¹ Code civil français, www.legifrance.gouv.fr/.

⁷⁰² Sullivan C. L. et Stalla-Bourdillon S., « Digital Identity and French Personality Rights – A Way Forward in Recognizing and Protecting an Individual's Rights in His/Her Digital Identity », *Computer Law & Security Review* (2015), www.ssrn.com/abstract=2584427.

⁷⁰³ *Mme X c. SARL Denim*, TGI de Paris, 17^{ème} chambre, décision du 21 novembre 2018.

Des exceptions à l'exigence de consentement peuvent exister lorsque les images sont prises dans un espace public, ou lorsqu'elles montrent des personnes connues dans le cadre de fonctions officielles ou d'autres activités liées à leur notoriété, sous réserve du respect du droit à la dignité des personnes. L'usage parodique et strictement privé figurent également au nombre des exemptions. Les tribunaux français peuvent également accorder des dommages et intérêts plus importants lorsqu'une célébrité a évité par le passé d'utiliser son image à des fins de parrainage ou d'autres exploitations commerciales. Le Code pénal prévoit également une infraction pour atteinte à la vie privée, assortie d'une sanction d'un an d'emprisonnement et d'une amende de 45 000 euros (Code pénal, article 226-1)⁷⁰⁴. Au-delà du domaine du droit au respect de la vie privée, le Code de la propriété intellectuelle français affirme qu'un artiste-interprète a le droit au respect de son nom, de sa qualité et de son interprétation (CPI, article 212-2)⁷⁰⁵. Dans la mesure où ce droit est imprescriptible et attaché à la personne, il est transmissible à ses héritiers pour la protection de l'interprétation et de la mémoire du défunt.

7.4.3. Suède

Pour un pays qui compte une population relativement faible, la Suède a donné naissance à un nombre remarquable de stars internationales de la télévision et du film, notamment Greta Garbo, Ingrid Bergman, Max von Sydow, Stellan Skarsgård, et Alicia Vikander. En plus d'être un haut lieu du genre du film noir scandinave, la Suède est à l'origine de plusieurs programmes qui ont connu un large succès dans d'autres marchés, tels que *Wallander*, *The Bridge*, ou la saga *Millenium*. Le fait que le pays ne dispose pas d'un droit de la personnalité distinct en tant que tel semble donc quelque peu surprenant.

Contrairement à ce qu'on observe en Allemagne ou en France, les droits de la personnalité sont pour ainsi dire absents du droit suédois⁷⁰⁶. Lorsque les médias d'information font un usage abusif d'une image, la loi sur la liberté de la presse s'applique, mais la Suède a tendance à faire prévaloir la liberté d'expression sur le droit à la vie privée des individus. Cette approche est quelque peu inhabituelle par rapport à celles de nombreux autres pays européens, et découle peut-être du fait que le droit suédois met au premier plan la transparence et l'ouverture, et donc également l'accès du public à l'information. Cette approche est illustrée par la *Utgivningsbevis*, un type controversé de licence de publication qui permet à des entreprises de publier ouvertement des informations personnelles relatives à des individus⁷⁰⁷. Dans le même ordre d'idée, le revenu

⁷⁰⁴ Code pénal français, www.legifrance.gouv.fr/.

⁷⁰⁵ CPI, formellement Code de la propriété intellectuelle, www.legifrance.gouv.fr/.

⁷⁰⁶ Ondreasova E., *Personality Rights in Different European Legal Systems : Privacy, Dignity, Honour and Reputation The Legal Protection of Personality Rights*, pp. 24–70. Brill | Nijhoff, https://brill.com/view/book/edcoll/9789004351714/B9789004351714_004.xml.

⁷⁰⁷ Herlin-Karnell E., « Corona and the Absence of a Real Constitutional Debate in Sweden », *Verfassungsblog on Matters Constitutional*, www.verfassungsblog.de/corona-and-the-absence-of-a-real-constitutional-debate-in-sweden/.

imposable des personnes relève du domaine public⁷⁰⁸. Ainsi, en Suède, le « principe de publicité » ne se réfère pas aux aspects des droits de la personnalité liés au droit à l'image, mais plutôt au principe selon lequel les documents officiels sont dans leur grande majorité non confidentiels et accessibles au public.

En dépit de ce qui précède, la loi suédoise relative aux noms et aux images en publicité impose l'obtention d'un consentement préalable à l'utilisation du nom ou de l'image de quelqu'un, ou d'une représentation qui identifie clairement un individu, pour des motifs commerciaux (Lag om namn och bild i reklam (1978:800))⁷⁰⁹. La loi suédoise sur la responsabilité délictuelle affirme de plus que les préjudices non financiers peuvent faire l'objet d'une compensation, mais uniquement lorsque l'infraction constitue un délit pénal ou qu'elle met en danger la santé ou la vie [Skadeståndslag (1972:207), chapitre 2]⁷¹⁰. Bien que la diffamation constitue une infraction pénale en vertu du Code pénal (Brottsbalken)⁷¹¹, elle n'est caractérisée que lorsque quelqu'un accuse une autre personne à tort ou sans « fondements raisonnables » d'être « un criminel ou de mener une vie répréhensible, ou qu'il rend publiques d'autres informations dans le but d'exposer au mépris d'autrui » (Brottsbalken, chapitre 5 §1). Le Code pénal permet également aux membres de la famille ou au procureur de lancer des poursuites pour « atteinte à la paix » du défunt, lorsque cela est dans l'intérêt public (Brottsbalken, chapitre 5 §4). En dehors de ces protections limitées, la législation suédoise en matière de droits de la personnalité présente des lacunes manifestes, et les juges suédois se sont montrés peu disposés à les combler par la voie jurisprudentielle⁷¹².

7.4.4. Guernesey

Le bailliage de Guernesey est une île située au large des côtes du nord de la France, et bien qu'il s'agisse d'une dépendance de la couronne britannique, elle dispose d'un système juridique distinct de celui du Royaume-Uni. Bien qu'elle ne compte que 63 000 habitants, son système fiscal avantageux pour les sociétés, ses paysages pittoresques et sa langue anglaise ont contribué à l'éclosion d'un secteur de la production cinématographique et du financement du film⁷¹³. Dans le contexte qui nous retient ici, il est intéressant de souligner que c'est Guernesey qui a mis en place le premier régime d'enregistrement du droit à l'image au monde, avec son ordonnance sur les droits à l'image de 2012⁷¹⁴. D'une façon

⁷⁰⁸ Marçal, K., « Sweden shows that pay transparency works », *The Financial Times*, www.ft.com/content/2a9274be-72aa-11e7-93ff-99f383b09ff9.

⁷⁰⁹ Lag om namn och bild i reklam (1978:800) [loi sur les noms et les images en publicité]. Traduction anglaise disponible sur : www.kb.se/Dokument/Bibliotek/biblioteksjuridik/names_pictures.pdf.

⁷¹⁰ Skadeståndslag (1972:207) [loi sur la responsabilité délictuelle suédoise]. Disponible sur : www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/skadestandslag-1972207_sfs-1972-207.

⁷¹¹ Brottsbalken, SFS 1962:700 [Code pénal suédois] Traduction anglaise disponible sur : www.legislationline.org/download/id/1700/file/4c405aed10fb48cc256dd3732d76.pdf.

⁷¹² Ondreasova E., *op.cit.*

⁷¹³ Tustin B., « Guernsey Film : More Than You Might Think », *Mondaq*, www.mondaq.com/guernsey/film-television/171220/guernsey-film-more-than-you-might-think.

⁷¹⁴ Ordonnance sur les droits à l'image (bailliage de Guernesey), 2012.

générale, la législation fonctionne de manière comparable aux lois qui s'appliquent à l'enregistrement des marques.

Les images protégées peuvent inclure les photos et représentations de l'individu, mais aussi des séquences de films, ainsi que son nom, sa voix, sa signature, son apparence, ses manières et attributs personnels, par exemple un numéro de maillot de sport. Les personnes vivantes, ou les personnes décédées dans les 100 ans suivant la demande, ainsi que les groupes, équipes, et même les personnages de fiction, peuvent être inclus au registre. Les demandeurs doivent détenir les droits à l'image en question : il peut donc s'agir de l'individu, ou de ses héritiers ou représentants agréés.

Une atteinte est caractérisée si l'image protégée enregistrée (ou une qui lui est similaire) est utilisée à des fins commerciales ou financières sans le consentement de son propriétaire et que cette utilisation soit source de confusion pour le public, ou bien qu'elle porte atteinte à la réputation de la personne représentée. Pour évaluer le préjudice subi, les tribunaux tiennent compte de tous les facteurs pertinents, notamment de toute conséquence économique ou perte de profits, ainsi que de tout préjudice moral subi par la victime. Les enregistrements sont valables pour une période initiale de dix ans, et peuvent être renouvelés pour dix années supplémentaires. Bien qu'elle ne soit pas liée à une obligation d'établissement ou de résidence à Guernesey, la mise en œuvre de ce droit sera limitée aux infractions et aux abus commis à Guernesey. Cependant, la loi a été élaborée avec à l'esprit les médias modernes et les services numériques transfrontières⁷¹⁵. Il serait donc prudent, quel que soit son lieu d'établissement ou de résidence, de procéder à des recherches d'antériorité avant de mettre des programmes à disposition sur l'île.

7.4.5. Royaume-Uni

Au moment où nous écrivons, le gouvernement en place au Royaume-Uni a souligné l'importance de désolidariser le pays de l'écosystème politique et économique de l'Union européenne. Mais malgré le Brexit, les films et émissions de télévision britanniques font partie intégrante du marché audiovisuel de l'Europe au sens large, et le public les apprécie partout sur le continent. En dépit de cela, le Royaume-Uni, comme la Suède, ne reconnaît pas formellement les droits de la personnalité dans sa législation. Et contrairement à bon nombre de leurs homologues européens, les tribunaux anglais se sont gardés de reconnaître ce droit au moyen de la jurisprudence.

Pour citer la décision dans l'affaire Rihanna mentionnée plus haut : « il n'existe pas aujourd'hui en Angleterre un droit autonome général pour une personne connue (ou toute autre personne) de contrôler la reproduction de son image ». Plutôt que de tenter de formuler un droit de la personnalité distinct, les juges du Royaume-Uni préfèrent habituellement s'appuyer sur les instruments juridiques plus traditionnels à leur disposition. Dans ce cadre, il existe plusieurs voies qui permettent à un individu de protéger

⁷¹⁵ Shires S (2015), « Guernsey image rights exposed », *Lexology*, www.lexology.com/library/detail.aspx?q=9829178f-e93d-4ead-94d0-bfbfaa81f8b7.

son image, contre des atteintes qui concernent généralement la divulgation de faits privés, et le préjudice porté à la réputation ou aux intérêts commerciaux d'une personne.

En raison de la façon dont la common law s'est développée au Royaume-Uni, et du fait que le pays ne dispose pas d'une constitution écrite, le droit à la vie privée n'est pas considéré comme un droit distinct. Les juges s'appuient donc plutôt sur le fait que le droit au respect de la vie privée constitue une valeur sociale importante qui sous-tend à son tour les lois spécifiques relatives à l'abus de confiance ou à l'usage abusif d'informations particulièrement sensibles. À partir de la fin des années 1990, les tribunaux ont abordé les protections de la vie privée dans le cadre de la loi du Royaume-Uni sur les droits de l'homme de 1998, qui reflète la Convention européenne des droits de l'homme. Cela étant, dans les affaires plutôt rares dans lesquelles les tribunaux anglais ont dû se pencher sur des questions de respect de la vie privée liées à des images, les juges ont généralement fait prévaloir la protection de la liberté d'expression et des autres libertés de la presse. Il existe des exceptions notables s'agissant des images qui présentent des informations confidentielles, ou sur lesquelles figurent des enfants ou des scènes à caractère particulièrement intime.

Lorsque le droit à l'image est protégé en tant qu'actif commercial, l'option la plus pertinente consiste souvent à s'appuyer sur le délit d'« imitation frauduleuse », mais le succès dépendra en premier lieu de l'importance de la « réputation ou de la notoriété » de la célébrité en question⁷¹⁶. Du point de vue de la dignité et de la réputation, un usage non autorisé de l'image d'une personne peut constituer en Common law le délit de mensonge préjudiciable, mais uniquement lorsqu'il consiste en des déclarations mensongères donnant lieu à un préjudice monétaire quantifiable⁷¹⁷. Les déclarations mensongères qui portent atteinte à la réputation d'un individu identifiable peuvent constituer une diffamation, mais depuis une réforme de la loi sur la diffamation de 2013⁷¹⁸, uniquement lorsqu'elles débouchent sur un « préjudice sérieux » pour l'individu représenté.

Ce qui précède doit également être lu à la lumière de la loi de 1988 relative à la propriété intellectuelle en matière de droit d'auteur, de modèles et de brevets (CDPA)⁷¹⁹ qui permet à un artiste-interprète, dans certaines circonstances, de « s'opposer à un traitement dérogatoire de la performance, par une distorsion, une mutilation ou d'autres modifications préjudiciables à la réputation de l'interprète » (CDPA, s. 205). Avant d'utiliser une performance enregistrée, un consentement doit être obtenu auprès de l'acteur, musicien, danseur ou autre interprète en question (CDPA, s. 182). Cette loi a été inspirée par une affaire portant sur des clips qui mettaient en scène l'acteur Peter Sellers, qui avaient été utilisés après son décès pour réaliser un nouveau film de la Panthère rose. Ses représentants successoraux étaient parvenus à faire valoir le droit post mortem de Sellers d'empêcher une exploitation non autorisée de ses performances⁷²⁰.

⁷¹⁶ *Edmund 'Eddie' Irvine c. Talksport* [2002] EWHC 367 (Ch).

⁷¹⁷ *Marathon Mutual Ltd c. Waters* [2009] EWHC 1931 (QB).

⁷¹⁸ Loi sur la diffamation de 2013, www.legislation.gov.uk/ukpga/2013/26/contents/enacted.

⁷¹⁹ Loi sur les dessins, les modèles et les brevets de 1988, www.legislation.gov.uk/ukpga/1988/48/contents.

⁷²⁰ *Rickless c. United Artists Corporation* [1988] QB 40.

7.4.6. Californie

Il n'existe pas de droit au respect de la vie privée au niveau fédéral aux États-Unis : les protections du droit à la vie privée sont réglementées dans le cadre de domaines ou secteurs spécifiques, comme les informations financières ou médicales par exemple. Bien qu'il existe des lois en matière de propriété intellectuelle et des codes relatifs à la concurrence et la publicité déloyales au niveau fédéral, ce sont avant tout les lois d'un État donné qui seront pertinentes en matière de protection de l'image et de la vie privée d'un individu. Parce qu'elle abrite à la fois Hollywood et la Silicon Valley, la Californie représente sans doute le mieux *l'esprit du temps* découlant de l'interaction entre célébrité et technologie.

La Californie protège le droit à l'image dans ses codes civils ainsi qu'en common law, en partie en raison de son industrie du divertissement de renommée mondiale et de la tendance californienne à tenir compte et à respecter les émotions de chacun. Le Code civil californien (CIV)⁷²¹ interdit l'usage non autorisé du nom, de la voix, signature, photographie ou représentation d'autrui à des fins publicitaires sans consentement (CIV §3344). Les contrevenants sont condamnés à verser à la partie lésée la somme la plus élevée entre 750 dollars américains (environ 635 euros en août 2020) ou le montant du préjudice effectivement subi, y compris une restitution de profit. La loi sur la protection de l'image des célébrités dite « Fred Astaire » a, par la suite, amendé le paragraphe §3344 afin de protéger l'utilisation commerciale du nom, de l'image ou de la voix d'une personne décédée pendant une durée de 70 ans après sa mort⁷²². Cependant, cette section s'applique uniquement à la marchandise, à la publicité et aux parrainages, et exemptent les œuvres fictionnelles ou non fictionnelles dramatiques, littéraires, musicales ou de divertissement. Le droit à l'image de Common law couvre tous les abus relatifs à l'identité d'un individu, ce qui va plus loin que les seules caractéristiques spécifiques figurant dans la loi, commis en vue de tirer des « bénéfices commerciaux ou autres »⁷²³.

Les tribunaux californiens reconnaissent que le droit d'une personne à être protégée d'une exposition non autorisée ou induite constitue un aspect du droit à la vie privée⁷²⁴. Plusieurs lois en matière de vie privée sont applicables à l'utilisation abusive de l'image de quelqu'un ; elles se distinguent selon qu'elles impliquent un préjudice de nature économique ou une atteinte à la dignité. Un contenu mensonger portant atteinte à la réputation de quelqu'un pourrait relever d'un délit de diffamation, et un contenu qui n'est pas techniquement mensonger mais qui porte néanmoins préjudice au bien-être mental ou émotionnel de la victime peut être constitutif d'un délit de « mise en lumière défavorable »⁷²⁵. En raison de la concentration de résidents qui se trouvent souvent sous le

⁷²¹ CIV (Code civil de Californie), <https://codes.findlaw.com/ca/civil-code/>

⁷²² Loi Fred Astaire sur la protection de l'image des célébrités, voir §3344.1 du Code civil californien.

⁷²³ *Clint Eastwood c. National Enquirer*, Superior Court, 149 Cal.App.3d 409 (Cal. Ct. App. 1983).

⁷²⁴ *Fairfield c. American Photocopy Equipment Co.*, (1955) 138 Cal. App. 2d 82, 291 P.2d 194

⁷²⁵ Dans les systèmes juridiques de Common law, une publication fausse qui nuit à la réputation d'une personne peut caractériser une diffamation. Mais dans certains États des États-Unis, dont la Californie, il existe un délit distinct (préjudice légal) qui couvre les communications qui ne sont pas techniquement fausses, mais qui sont néanmoins trompeuses. À cet égard, les tribunaux de Californie reconnaissent que de fausses *implications*

feu des projecteurs (tant métaphoriquement que littéralement), le système juridique californien est considéré comme le plus favorable aux requérants qui souhaitent lancer une procédure en violation des droits de la personnalité.

Il serait toutefois négligent de ne pas souligner que les protections concurrentes en matière de liberté d'expression découlant du premier amendement sont robustes, même s'agissant d'images et de vidéos publiées par des entreprises. Un exemple notable concerne une procédure lancée par l'acteur Dustin Hoffman contre un magazine qui avait manipulé numériquement une image de l'acteur pour illustrer un sujet de mode en le montrant habillé de vêtements de créateur. Les tribunaux ont conclu que dans la mesure où le magazine n'avait pas l'intention de porter un tort à l'acteur et que l'image en cause n'avait pas été réalisée purement pour des motifs commerciaux, son éditeur pouvait se prévaloir de la liberté d'expression⁷²⁶.

Les personnes connues, y compris les individus qui se retrouvent impliqués dans des événements dignes d'intérêt public, disposeront généralement d'une moindre protection de la vie privée que des membres du public ordinaires. De façon significative pour ceux qui chercheraient à modifier l'apparence d'un acteur au moyen de techniques de cinéma fantôme ou autres, la Californie s'appuie sur un test du caractère transformatif d'une œuvre pour déterminer si l'utilisation de l'image d'une personne est protégée par le premier amendement. Selon ce test, plus une nouvelle œuvre « transforme » les prises de vue d'origine pour proposer un sens ou un message différent, plus il est probable qu'elle ne constitue pas une infraction au droit d'auteur (Code sur le droit d'auteur, §107)⁷²⁷. S'agissant plus précisément des images d'individus, les tribunaux californiens examineront si « la représentation d'une célébrité est transformée au point de constituer principalement l'expression propre du défendeur plutôt que la représentation de la célébrité »⁷²⁸.

7.5. Quel avenir pour le secteur audiovisuel européen ?

Ce chapitre visait à exposer la nature multiforme des droits de la personnalité, qui comprennent généralement divers droits en matière d'image et de respect de la vie privée. Il convient également d'intégrer à ce cadre les droits à la dignité et à l'intégrité, ainsi que les droits voisins des artistes-interprètes. Mais quelle que soit la façon dont ces droits sont

peuvent entraîner des impressions fausses, et au bout du compte préjudiciables, vis-à-vis de l'individu concerné. Cette distinction est subtile et nuancée, mais en général, la diffamation cherche à corriger le préjudice porté à la réputation d'une personne. La vérité permet de se défendre contre une action en diffamation, c'est-à-dire que si la déclaration est vraie, il n'y a pas diffamation. Mais une photographie ou une déclaration factuellement correcte peut « montrer quelqu'un sous une fausse lumière ». Le délit californien de « mise en lumière défavorable » se préoccupe donc du préjudice porté aux émotions d'une personne – sans tenir compte de la véracité. CACI (Judicial Council of California Civil Jury Instructions) édition 2017. N° 1802. False Light. <https://www.justia.com/trials-litigation/docs/caci/1800/1802/>.

⁷²⁶ *Dustin Hoffman c. Capital Cities/ABC, Inc.*, 255 F. 3d 1180 (9^{ème} Cir. 2001).

⁷²⁷ Code sur le droit d'auteur, formellement U.S. Code, Title 17 Copyrights, §107 Limitations on exclusive rights: Fair use, www.law.cornell.edu/uscode/text/17/107.

⁷²⁸ *Comedy III Prods., Inc. c. Gary Saderup, Inc.* - 25 Cal. 4th 387, 106 Cal. Rptr. 2d 126, 21 P.3d 797 (2001).

classifiés du point de vue théorique, certaines observations et approches peuvent aider à atténuer les risques associés aux *deepfakes* et à l'utilisation d'acteurs fantômes.

Certains pays ont adopté ou proposé des lois spécifiques aux *deepfakes* pour traiter le problème de l'abus sexuel au moyen d'images ou de la question de l'interférence dans les processus électoraux. Par exemple, l'État de Virginie aux États-Unis a été le premier à réviser ses lois en matière de « pornodivulgateur » pour y inclure un délit de publication de photos ou vidéos manipulées montrant des personnes nues ou leurs parties intimes sans consentement⁷²⁹. Le Texas a été le premier à criminaliser le partage de vidéos de *deepfakes* réalisées dans le but de causer du tort à un candidat dans les jours précédant une élection⁷³⁰. Bien entendu, les *deepfakes* peuvent être profondément dommageables sans être nécessairement de nature sexuelle ou politique.

À défaut d'instruments juridiques, les grandes entreprises du secteur technologique ont tenté de réguler les *deepfakes* apparaissant sur leurs services. Facebook et sa société sœur Instagram ont officiellement « interdit » les *deepfakes*, tout comme l'ont fait Twitter, Reddit et Pornhub. Bien que TikTok ait par le passé plutôt bien accueilli cette pratique, la société l'a également récemment « interdite » à la suite de discussions portant sur une prise de contrôle par des sociétés américaines⁷³¹. Le recours aux guillemets autour du mot « interdite » est intentionnel : en dépit des politiques et déclarations officielles, le retrait et la modération des *deepfakes* sont d'une extrême complexité, aussi leur interdiction n'est-elle souvent que théorique.

Premièrement, distinguer un *deepfake* d'une vidéo authentique présente un défi considérable en soi, comme l'a montré l'expérience du défi de détection des *deepfakes* mené par Facebook, Microsoft et un bataillon d'experts de l'intelligence artificielle⁷³². Certains efforts portent sur le méta-étiquetage des images sources et le recours aux filigranes, mais il n'existe aucune solution largement accessible à ce jour. Deuxièmement, même si un *deepfake* était détecté, il est potentiellement impossible d'identifier la personne qui en est à l'origine, au vu de la facilité et de l'opportunité de conserver son anonymat en ligne. En plus de ces difficultés techniques claires, des complications liées à l'intention et au contexte doivent également être prises en compte. Le terme de « *deepfakes* » se réfère à une méthode de création et notamment à la substitution de visages, mais non à leur substance : la médecine et l'éducation sont deux domaines dans lesquels cette technologie a été admirablement exploitée dans l'intérêt de la société⁷³³. En bref, il n'y a pas toujours

⁷²⁹ Loi 2678 de la chambre des délégués de Virginie, formellement Bill to amend and reenact § 18.2-386.2 of the Code of Virginia, relating to unlawful dissemination or sale of images of another ; falsely created videographic or still image,.

<https://law.lis.virginia.gov/vacode/title18.2/chapter8/section18.2-386.2/#~:text=Any%20person%20who%2C%20with%20the,or%20female%20breast%2C%20where%20such.>

⁷³⁰ Loi 751 du Sénat du Texas, formellement Bill relating to the creation of a criminal offense for fabricating a deceptive video with intent to influence the outcome of an election, www.legiscan.com/TX/text/SB751/2019.

⁷³¹ Statt N., « TikTok is banning deepfakes to better protect against misinformation », *The Verge*, www.theverge.com/2020/8/5/21354829/tiktok-deepfakes-ban-misinformation-us-2020-election-interference.

⁷³² Wiggers K., « Facebook, Microsoft, and others launch Deepfake Detection Challenge », *VentureBeat*.

⁷³³ Kalmykov M., « Positive Applications for Deepfake Technology », *Hackernoon.com*, www.hackernoon.com/the-light-side-of-deepfakes-how-the-technology-can-be-used-for-good-4hr32pp.

une ligne de démarcation claire entre *deepfakes* indésirables d'un côté et acceptables de l'autre.

Au-delà de la manière dont les « mauvais » *deepfakes* pourraient être séparés des « bons », il faut noter que les contenus novateurs et divertissants ont une incroyable capacité à devenir viraux, même s'ils sont faux ou trompeurs. Bien qu'il s'agisse là d'un point de vue assez cynique, les plateformes qui dépendent du nombre de clics publicitaires ou de pages vues pourraient estimer qu'il est dans leur intérêt commercial de garder les contenus les plus attrayants en ligne. De plus, même lorsqu'il convient vraisemblablement de retirer un certain *deepfake*, une telle décision peut avoir un effet dissuasif en matière de liberté d'expression. Enfin, même lorsqu'un *deepfake* a été retiré d'une plateforme, il peut s'agir d'une solution de surface, dans la mesure où la vidéo pourrait facilement réapparaître autre part. De plus, l'effet dit Streisand⁷³⁴ a démontré qu'une tentative de suppression d'une information pouvait involontairement aboutir à élargir sa diffusion.

Lorsque des contenus audiovisuels sont publiés en ligne, il peut être difficile de les protéger d'un usage non souhaité, ou d'un détournement par des parties tierces. Il est néanmoins toujours possible à des parties privées de s'accorder par voie contractuelle sur la façon dont les droits de la personnalité pourront être exploités ou protégés. Les interprètes concluent généralement un contrat nouveau pour chacun de leurs projets, et la définition de paramètres clairs dès le départ pourrait être utile pour prévenir des litiges ultérieurs. Il n'est pas besoin de recourir à un jargon juridique pour ce faire, mais il est prudent de préciser par écrit les façons dont l'image d'une personne pourra être utilisée, partagée ou transférée à d'autres.

Les deux parties auraient tout intérêt à examiner minutieusement la manière dont le contenu pourrait être utilisé au-delà de la production initiale, ainsi que la question d'une rémunération complémentaire. Lorsqu'un studio réutilise les images d'une performance précédente pour une production nouvelle, il est peu probable que cela donne lieu à des paiements supplémentaires si ceux-ci ne sont pas clairement stipulés dans le contrat. Du point de vue de la réputation, il pourrait également être utile d'inclure des clauses anti-dénigrement ou des clauses dites de moralité, qui visent à empêcher une partie de nuire à la réputation de l'autre. Pour les acteurs, celles-ci seront d'autant plus importantes dans des pays tels que le Royaume-Uni, la Suède ou les États-Unis, dans lesquels la diffamation et l'atteinte à la dignité sont notoirement difficiles à faire valoir.

D'un point de vue strictement financier, une modification indésirable des apparences de l'acteur pourrait nuire, ou même détruire, sa relation avec ceux qui travaillent derrière la caméra. Les demandes et les différends créatifs, qu'ils soient ou non raisonnables, sont connus pour faire dérailler ou même couler des productions. La perte ou

⁷³⁴ L'effet Streisand tire son nom de l'action en justice intentée en 2003 par l'actrice et chanteuse américaine Barbra Streisand contre un photographe, Kenneth Adelman. Streisand tentait de faire interdire la publication des photographies qu'Adelman avait prises de sa maison située sur les plages californiennes de Malibu. Cependant, la photographie d'Adelman avait été réalisée dans le cadre du California Coastal Records Project, un projet de recherche scientifique portant sur l'érosion des côtes qui met à disposition de la recherche des photos de la côte californienne. La demande de l'actrice a fini par être rejetée mais la publicité qu'avait générée le procès a encore accru l'intérêt public pour les photos de sa maison. Voir Cacciottolo M., « The Streisand Effect : When censorship backfires », *BBC News*, <https://www.bbc.co.uk/news/uk-18458567>.

le remplacement d'une star en cours de tournage peut coûter des milliers ou des millions d'euros à un studio. Des modifications substantielles, au-delà des besoins artistiques du film, devraient donc être discutées en amont, en particulier lorsque ces altérations sont susceptibles de porter préjudice aux intérêts ou aux sentiments de l'acteur.

Enfin, bien qu'il puisse être fascinant et innovant de créer le double numérique ou une version fantôme de l'image de quelqu'un, cela présente un risque inévitable. Pour paraphraser Ingmar Bergman, lorsque nous nous asseyons pour profiter d'un film ou d'une émission de télévision, nous nous préparons de façon consciente à une illusion. Nous mettons de côté la volonté et l'intellect de façon à laisser un récit fictif se déployer dans notre imaginaire⁷³⁵. Mais il s'agit de tout autre chose lorsqu'on considère le réalisme des visages créés désormais, parce qu'il pourrait porter atteinte à l'autonomie et à l'autodétermination des acteurs concernés. Ce réalisme a également tendance à manipuler les processus d'évaluation et de prise de décision du public. Quand les individus sont représentés de manière mensongère dans des vidéos non fictionnelles, comme des documentaires, des interviews ou des publicités, le risque de préjudice financier, réputationnel et sociétal est encore plus manifeste. On peut en toute logique s'attendre, au fur et à mesure que l'intelligence artificielle devient de plus en plus sophistiquée, à ce que la distinction entre performances authentiques et celles générées numériquement se complexifie. Si l'œil humain est incapable de discerner la différence, la loi cessera peut-être également de distinguer les deux et étendra ainsi les droits de la personnalité pour couvrir l'identité virtuelle des personnes.

S'il fallait retenir une seule chose de ce chapitre, c'est que les droits de la personnalité nécessitent une prise en compte minutieuse du contexte. Lorsque ces droits font l'objet de procédures judiciaires, il revient aux tribunaux de les interpréter plutôt que de créer du droit, et donc, d'affirmer simplement l'applicabilité de règles établies qui ont évolué par rapport aux usages de cette société. Il n'est pas simple, même pour quelqu'un qui bénéficie d'une formation ou d'une autorité juridique, de trancher des questions relatives à l'art, la vérité, l'expression, l'identité – qui touchent toutes à l'essence de ce que signifie « être humain ». Lorsqu'on s'interroge sur l'exploitation et la protection des droits de la personnalité, en particulier en ce qui concerne les technologies novatrices telles que les *deepfakes* et le cinéma fantôme, il semble qu'il faille se contenter de la réponse habituelle de l'avocat, du moins pour le moment : « ça dépend ».

⁷³⁵ Bergman I., *Four Screenplays*, Secker & Warburg, London. Traduction du suédois vers l'anglais par L. Malmstrom et D. Kushner.



La réglementation de l'intelligence artificielle

*Le principe d'une « prise en compte minutieuse du contexte » mentionné par Kelsey Farish dans sa contribution à cette publication devrait probablement être étendu à l'ensemble des questions juridiques abordées dans ces pages. Et il faut garder à l'esprit ce fait central (déjà évoqué à l'avant-propos de cette publication) : les ordinateurs restent des ordinateurs, des machines stupides capables en tout et pour tout de distinguer un 1 d'un zéro. Ainsi, les résultats de leurs efforts de calculs sans âme sont tributaires et/ou auront besoin d'une intervention et d'un contrôle humains. Et bien souvent, une intervention humaine signifie une réglementation. **Atte Jääskeläinen** présente dans sa contribution à cette publication les principes qui devraient selon lui être retenus pour une réglementation de l'IA. Comme mentionné plus haut, la transparence constitue dans ce contexte le principe le plus fondamental, puisqu'elle « sert les besoins humains pour donner un sens au fonctionnement des systèmes et confier des responsabilités aux bonnes personnes ». Selon Jääskeläinen, nous devrions néanmoins accepter que « des risques inconnus puissent être impossibles à réglementer, du moins si la réglementation est basée sur la technologie et non sur des objectifs ». De plus, la réglementation de l'IA devrait « réduire le risque public sans détruire la créativité et l'innovation, et les obstacles inutiles à l'utilisation des données pour faire le bien et créer du bien-être » devraient être éliminés.*

8. Approches pour un cadre réglementaire durable pour les industries audiovisuelles en Europe

Atte Jääskeläinen, Université LUT, Finlande, et LSE, Londres⁷³⁶

8.1. Introduction

Un logiciel repère le ballon et les joueurs dans l'image d'un match de football et effectue un zoom sur l'endroit où se déroule l'action. Et comme l'image des caméras est très nette, le « réalisateur virtuel », un ordinateur qui s'appuie sur des algorithmes d'apprentissage automatique, est en mesure de proposer un match entier avec plusieurs angles de caméra - sans intervention humaine. L'objectif de la KNVB, la fédération néerlandaise de football, et de la société de médias néerlandaise Talpa est de diffuser 80 000 matchs de football amateur en direct chaque année, sans aucune intervention humaine.

Un logiciel analyse votre comportement dans les médias sociaux et peut identifier votre personnalité sur la base du modèle psychologique largement utilisé, Big 5, atteignant la même précision que celle obtenue en complétant un questionnaire avec des informations sur quelque 200 likes de Facebook. Le résultat : un système qui pourrait être utilisé pour influencer les élections en adressant les messages politiques aux personnes les plus susceptibles d'être convaincues par un certain style de publicité personnalisée.

Un logiciel développé par l'éditeur régional suédois Mittmedia combine la personnalisation avec la fixation d'objectifs journalistiques, selon une approche claire de contextualisation des données et d'utilisation de l'apprentissage automatique. Le directeur des techniques informatiques de la société prévoit une personnalisation et une automatisation totales de l'édition dans un avenir proche.

⁷³⁶ Atte Jääskeläinen est professeur de pratique à l'Université LUT, en Finlande, et chercheur invité à la LSE, à Londres. Il a été directeur de l'information et des affaires courantes de la société de radiodiffusion finlandaise de 2006 à 2017 et PDG de l'agence de presse finlandaise de 2004 à 2006. Il est co-auteur avec Maike Olij du rapport d'actualités de l'UER 2019 « The Next Newsroom. Unlocking the Power of AI for Public Service Journalism ».

Un logiciel développé par la société suisse Largo prédit, à partir d'un premier scénario, la réaction du public, pays par pays. Sur la base de cette analyse, les spécialistes des données indiquent quelles parties du scénario devraient être modifiées pour augmenter l'attrait commercial d'un film et maximiser les revenus générés.

Tous ces exemples concrets montrent comment l'intelligence artificielle affecte déjà l'industrie audiovisuelle ou son environnement de manière stratégique : la production audiovisuelle peut être largement automatisée ; le comportement du public peut être analysé comme jamais auparavant ; ces informations peuvent être utilisées à la fois pour optimiser la valeur que les clients tirent du marché, mais aussi pour guider le travail créatif et journalistique et l'utiliser dans les processus démocratiques - parfois d'une manière considérée comme incompatible avec la démocratie.

Pour certains, cette évolution conduira l'humanité à une catastrophe, si elle n'est pas correctement réglementée. Comme l'a écrit l'Amu Webb de l'Institut Future Today en 2019 : « Le manque de nuance est une partie du problème de la genèse de l'IA : certains surestiment considérablement l'applicabilité de l'IA dans leur milieu professionnel, tandis que d'autres voient en elle une arme imparable »⁷³⁷.

La réglementation de l'IA, en particulier dans les domaines qui concernent l'industrie audiovisuelle, est une question complexe sans réponse claire. Cette complexité affecte la prise de décision sur le terrain. Pour les autorités et les politiciens, le domaine de l'IA est difficile à cerner, car le concept de l'IA est flou et la cible évolue rapidement : lorsque l'on pense l'avoir enfin saisie, de nouvelles formes ont déjà vu le jour. La réglementation est un processus réactif et, dans un domaine qui évolue rapidement, elle devient souvent obsolète avant de pouvoir être appliquée⁷³⁸.

Certains des exemples d'IA donnent l'impression que les ordinateurs créent quelque chose de presque magique. En général, cette illusion repose simplement sur l'utilisation de mathématiques intelligentes rendue possible par les énormes ressources de calcul disponibles grâce à l'informatique en nuage et aux télécommunications rapides. Ce qui apparaît comme de la créativité est simplement l'absence de préjugés qui limitent la pensée humaine, combinée à la capacité de vérifier et de produire toutes les alternatives, y compris celles que les humains n'envisageraient pas même lorsqu'ils sont appelés à penser hors des sentiers battus.

Les auteurs d'un livre récent sur « l'économie simple » de l'IA prédictive expliquent : « Toutes les activités humaines peuvent être décrites par cinq composantes de haut niveau : les données, la prédiction, le jugement, l'action et les résultats. À mesure que l'intelligence des machines s'améliorera, la valeur des compétences humaines en matière de prédiction diminuera, car la prédiction des machines constituera un substitut moins coûteux et plus

⁷³⁷ Webb A., « The big nine before it's too late », *WMG Weekly*, <https://www.wmgweekly.com/post/2019/06/08/the-big-ninebefore-it-s-too-late>.

⁷³⁸ Petit N., « Law and regulation of artificial intelligence and robots - Conceptual framework and normative implications », https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339.

efficace. Cependant, la valeur des compétences humaines en matière de jugement augmentera.»⁷³⁹

L'OCDE a estimé que près de la moitié des professions vont soit disparaître, soit changer fondamentalement dans les 15 à 20 ans en raison de l'automatisation et des nouvelles technologies d'auto-apprentissage⁷⁴⁰. D'un point de vue pratique, il semble que le défi consiste à aligner les processus de telle sorte que la force à la fois des machines et des hommes puisse être utilisée de la meilleure manière possible.

Les technologies n'existent pas isolément de la culture et des valeurs. Certains universitaires considèrent que l'automatisation croissante entraîne une redéfinition majeure des valeurs en lien plus étroit avec l'informatique⁷⁴¹. D'autres affirment que la communication ne peut plus être définie comme un phénomène n'existant que de personne à personne⁷⁴². Une observation fondamentale est que nous sommes en train de passer à une ère de communications multiples, dans laquelle il existe des milliards et des milliards de connexions entre les individus qui sont alimentées par l'automatisation et définissent notre compréhension du monde en réseau dans lequel nous vivons⁷⁴³.

Ce chapitre se propose de clarifier certaines des questions éthiques et morales qui devraient être prises en compte lors de l'élaboration d'une réglementation efficace et éthiquement saine pour ce domaine technologique qui est à la fois fascinant et, pour certains, effrayant.

8.1.1. Les bases, simplifiées, de l'IA

L'expression « intelligence artificielle » n'est pas claire. Une définition courante veut que l'IA décrive des processus automatiques qui nécessiteraient de l'intelligence s'ils étaient réalisés par des humains. L'expression a été inventée à l'origine par John McCarthy, qui a commencé ses recherches sur l'IA dans les années 1950. Il supposait que l'apprentissage et l'intelligence de l'homme pouvaient être simulés par une machine. Mais dans sa forme la plus élémentaire, l'intelligence artificielle est « un système qui prend des décisions autonomes, une branche de l'informatique dans laquelle les ordinateurs sont programmés pour faire des choses qui requièrent normalement de l'intelligence humaine »⁷⁴⁴. Un des

⁷³⁹ Agrawal A., Gans J. et Goldfarb A., « Prediction machines: The simple economics of artificial intelligence », Harvard Business Press.

⁷⁴⁰ OCDE, L'avenir du travail. Perspectives de l'emploi de l'OCDE 2019. Faits marquants, https://www.oecd-ilibrary.org/fr/employment/perspectives-de-l-emploi-de-l-ocde-2019_b7e9e205-fr.

⁷⁴¹ Coddington M., *Clarifying journalism's quantitative turn*, Digital Journalism, Routledge, 3(3), pp. 331–348.
Milosavljević M. et Vobić I., 'Our task is to demystify fears': *Analysing newsroom management of automation in journalism*, Journalism, SAGE Publications, <https://journals.sagepub.com/doi/abs/10.1177/1464884919861598?journalCode=joua>.

⁷⁴² Lewis S. C., Guzman A. L. et Schmidt T. R., *Automation, journalism, and human-machine communication: Rethinking roles and relationships of humans and machines in news*, Digital Journalism, Routledge, 7(4), pp. 409–427.

⁷⁴³ Jääskeläinen A. et Olij M., « The next newsroom: Unlocking the power of AI for public service journalism », Union européenne de radio-télévision, <https://www.ebu.ch/publications/news-report-2019>.

⁷⁴⁴ <https://futuretodayinstitute.com/trend/artificial-intelligence/>.

pionniers de l'intelligence artificielle, Marvin Minsky, a décrit l'intelligence artificielle comme un « terme fourre-tout » : elle couvre de nombreux concepts.

La plupart des applications de l'intelligence artificielle utilisent aujourd'hui des technologies qui relèvent du domaine de l'apprentissage automatique ou machine (*machine learning*). Avec l'apprentissage automatique, les ordinateurs apprennent à partir de données sans être explicitement programmés.

- Dans la programmation traditionnelle, les données sont traitées par un ordinateur qui produit un résultat. Les programmeurs doivent donc connaître les règles pour pouvoir obtenir le résultat souhaité.
- Dans l'apprentissage automatique, les données et des exemples du résultat souhaité sont d'abord traités par l'ordinateur qui en tire ensuite des enseignements pour créer ses propres règles. Ces règles sont ensuite utilisées pour produire le résultat.

Une forme d'apprentissage machine très puissante et consommatrice de ressources, l'apprentissage profond (*deep learning*), utilise des algorithmes appelés « réseaux de neurones artificiels », qui sont calqués sur la façon dont les réseaux de neurones fonctionnent dans le cerveau humain. Les progrès de l'apprentissage profond ont rendu les technologies du langage et la reconnaissance d'images beaucoup plus sophistiquées et ont ainsi fait naître de nouvelles possibilités dans le secteur de l'audiovisuel.

Le principal avantage de l'apprentissage profond est qu'il est capable d'absorber d'énormes quantités de données. Cette capacité a permis à l'apprentissage automatique d'accomplir des tâches qu'il n'aurait jamais pu gérer auparavant. D'autre part, il nécessite également d'énormes quantités de données de formation et des ressources informatiques coûteuses, de sorte que son utilisation à grande échelle n'est devenue possible que récemment - et reste encore limitée.

Les différents types d'apprentissage automatique sont basés sur la façon dont les machines utilisent les données pour apprendre des règles. Le plus ancien type d'apprentissage automatique est appelé apprentissage supervisé. Il utilise un ensemble de résultats souhaités pour entraîner l'ordinateur. L'algorithme propose alors des règles qui permettront à l'ordinateur de produire des résultats similaires à ceux de l'ensemble des données de formation. Dans l'apprentissage non supervisé, l'ordinateur est utilisé pour regrouper un énorme ensemble de données de manière significative. Cette approche n'exige pas un ensemble de données de formation, car l'ensemble de données est généralement regroupé en fonction de sa logique interne.

L'un des modèles les plus récents est l'apprentissage par renforcement, dans lequel le système apprend à la volée à partir des rétroactions qu'il reçoit de son environnement. Ce domaine se développe rapidement et est très utile car les systèmes basés sur l'apprentissage par renforcement peuvent rapidement s'adapter aux nouvelles situations et tirer des enseignements du comportement des utilisateurs – qui peut changer soudainement, pour une raison ou pour une autre. Par exemple, les systèmes de recommandation les plus avancés utilisent l'apprentissage par renforcement comme moyen de contrôler et de s'adapter au comportement des utilisateurs.

Ces trois modèles d'apprentissage peuvent être combinés en un système complexe. Un système avancé possède généralement plusieurs algorithmes et peut combiner différentes approches pour obtenir le résultat souhaité dans une situation spécifique. Pour ce type de système, il est crucial de comprendre le problème, le contexte et les informations contenues dans les données.

Ce qui porte à confusion avec les différents domaines de l'intelligence artificielle, c'est que les gens confondent souvent les dimensions de l'IA. Par exemple, l'apprentissage automatique apparaît souvent à côté de la reconnaissance d'images et du traitement du langage naturel dans les listes des principaux domaines d'application de l'IA. Mais l'apprentissage automatique est présent dans presque toutes les applications modernes de l'intelligence artificielle, y compris le diagnostic médical, les voitures autonomes, les systèmes de prédiction et la classification automatique.

Concernant la nécessité de réglementer l'IA, une catégorisation très importante est l'IA étroite ou faible, dans laquelle le système effectue une seule tâche liée à un problème spécifique. Un domaine de réglementation beaucoup plus difficile est celui de l'intelligence artificielle générale (IAG), qui est un type d'IA capable de résoudre des problèmes complexes dans n'importe quel contexte et de définir ses objectifs de manière autonome. Bien qu'il existe des initiatives audacieuses pour parvenir à ce type d'autonomie des ordinateurs, comme DeepMind et OpenAI de Google, ce type d'IA n'est pas encore opérationnel. La plupart des chercheurs pensent que nous n'atteindrons pas ce niveau d'IA avant, au moins, plusieurs dizaines d'années⁷⁴⁵.

8.2. Comment l'IA est-elle utilisée dans les industries audiovisuelles ?

Par industries audiovisuelles, nous entendons traditionnellement la production et la commercialisation des films, des émissions de télévision et du contenu audiovisuel sur internet. Cependant, comme il est difficile de définir des limites claires pour l'IA, il n'est pas sage de limiter notre réflexion aux seules spécificités de ce secteur particulier. De nos jours, la diffusion de messages se fait souvent par des méthodes mixtes – et ces méthodes peuvent même être sélectionnées de manière autonome, en fonction des préférences individuelles, par des systèmes d'IA dirigés par les données. Par conséquent, les défis de la réglementation de l'IA dans le secteur de l'audiovisuel sont fortement liés à d'autres secteurs connexes. La question clé est la suivante : qu'essayons-nous de réaliser par la réglementation et comment devrait-elle être appliquée dans les domaines pertinents pour ce secteur ? L'éventail des applications potentielles est déjà diversifié et se développe rapidement dans des domaines que nous ne pouvons même pas encore imaginer. Toutefois,

⁷⁴⁵ Voir, par exemple, Joshi N., *et al.* « How far are we from achieving artificial general intelligence? », <https://www.forbes.com/sites/cognitiveworld/2019/06/10/how-far-are-we-from-achieving-artificial-general-intelligence/#5ebe24f06dc4> et Fjelland R., « Why general artificial intelligence will not be realized », <https://www.nature.com/articles/s41599-020-0494-4>.

il ne serait ni sage ni efficace d'appliquer les mêmes règles aux traductions automatiques, aux voitures autonomes, aux données à caractère personnel sensibles et aux technologies avancées de caméras, pour ne citer que quelques-uns de ces domaines d'application.

Comment aborder alors cette question ? Premièrement, nous devrions essayer de mieux comprendre les conséquences des différents usages de l'IA dans les industries audiovisuelles. Comme elles varient sensiblement, même une catégorisation grossière aide à comprendre les valeurs qui devraient être protégées par la réglementation.

En l'absence d'une meilleure catégorisation, plus sophistiquée, on peut utiliser ici une catégorisation employée dans le rapport pour l'Union européenne de radio-télévision sur la façon d'utiliser l'IA dans le journalisme de service public, « The Next Newsroom »⁷⁴⁶, qui pourrait servir de cadre de base pour comprendre la pertinence stratégique des différents types de technologies d'IA pour l'industrie audiovisuelle.

Premièrement, l'intelligence artificielle peut être considérée comme un ensemble grandissant d'outils pratiques. Il s'agit de l'IA à un niveau purement opérationnel, visant principalement à automatiser les tâches répétitives et à réduire les coûts.

Par exemple, les solutions comprenant des systèmes d'IA pour les tâches de montage et de gestion des médias sont nombreuses, et leur adoption progresse rapidement. Les outils utilisés pour la transcription et la traduction, et pour la détection de matériel spécifique dans les archives, permettent de réutiliser le matériel plus facilement et plus rapidement, et modifient la logique de création de valeur dans la production. Par exemple, Deutsche Welle utilise un traitement linguistique basé sur l'IA pour que la salle de rédaction puisse suivre en temps réel ce qui se passe dans ses services d'information dans plusieurs langues. Swiss Broadcasting est l'une des sociétés qui a développé des systèmes avancés pour détecter les personnes et les lieux dans les archives vidéo⁷⁴⁷.

Le radiodiffuseur irlandais RTE et Al Jazeera ont même collaboré pour créer un système qui mesure le temps d'antenne des hommes politiques pendant les campagnes électorales et signale les contenus susceptibles de poser un problème de réglementation. Le système est basé sur une méthode avancée de détection non seulement des objets et des personnes dans les images, mais aussi de leur contexte⁷⁴⁸.

On peut raisonnablement se poser les questions suivantes : Y a-t-il quelque chose de spécial dans ces outils qui nécessite une réglementation spécifique ? Ou suffit-il simplement que ces outils soient utilisés de manière responsable ? Après tout, il ne s'agit que d'outils et de technologies qui nous aident à accomplir notre travail.

Deuxièmement, l'IA rend possible la création d'une culture sachant analyser les données, qui repose sur la définition et la connaissance de vos objectifs et l'apprentissage de moyens de mesure et d'optimisation, basés sur ces objectifs. Cela permet également une utilisation très stratégique de l'IA : ciblage des messages et optimisation de la valeur pour

⁷⁴⁶ Jääskeläinen A. et Olij M., *op.cit.*

⁷⁴⁷ Rezzonico P., « Artificial intelligence at the service of the RTS audiovisual archives », FIAT/IFTA, <http://fiatifta.org/index.php/artificial-intelligence-at-the-service-of-the-rtt-audiovisual-archives/>

⁷⁴⁸ TM Forum, « AI indexing for regulatory practice », <https://www.tmforum.org/ai-indexing-regulatory-practise>.

les clients individuels en fonction d'informations sur leurs préférences et leur comportement.

Le même type d'IA d'optimisation et de personnalisation est également utilisé pour optimiser les résultats financiers des opérations audiovisuelles, par exemple en rendant les campagnes de marketing plus efficaces grâce au test de l'efficacité de différents messages ou à l'identification des opportunités et des groupes de marché intéressants.

À l'ère de l'abondance que les technologies numériques ont créée, l'un des défis fondamentaux consiste à trouver le bon contenu dans un trop grand fouillis. Il est donc nécessaire d'établir un lien entre le contenu et le public qui s'y intéresse. Ce type d'optimisation ne se limite pas aux offres en ligne. Par exemple, la RTVE espagnole a un projet de recherche prometteur sur l'élaboration d'une grille de programmes télévisuelle à l'aide d'algorithmes d'IA. La question posée est la suivante : quels programmes de télévision correspondent au goût du public à une heure précise de la journée ou un jour précis de la semaine ?⁷⁴⁹

Il est déjà possible techniquement de connecter tous les appareils d'un utilisateur en lui soumettant des informations sur ses préférences musicales, les films qu'il a regardés, ses habitudes télévisuelles et même de détecter son humeur à partir de ses appareils de santé personnels, et de le diriger ensuite vers le meilleur contenu. Ajoutez des informations sur la météo, votre agenda professionnel et les conditions de circulation locales, et vous aurez un assistant personnel assez puissant qui vous proposera ce qui peut vous intéresser à ce moment précis, à l'endroit où vous vous trouvez. Tous ces services existent déjà, et la plupart d'entre eux sont basés sur des contenus audiovisuels. En réalité, certains des géants technologiques mondiaux proposent déjà quelque chose de ce genre. Il suffit de penser à ce que l'assistant Google ou le portefeuille de services d'Apple sont capables de faire. Toutes ces fonctionnalités sont combinées sur le système d'exploitation de notre téléphone mobile – contrôlé par ces mêmes sociétés.

Troisièmement, l'IA peut être utilisée dans des processus uniques visant à créer un meilleur contenu, plus distinctif. Cela implique non seulement d'optimiser et de répéter ce qui a déjà été fait, mais aussi de créer des approches totalement nouvelles sans les limites et les biais du cerveau humain. Ce domaine pose des défis intéressants car parfois, un bon résultat créé par une machine n'est pas vraiment meilleur que celui créé par un humain, ou peut ne pas être acceptable selon notre code éthique. Un autre angle intéressant est la question de savoir comment maintenir la créativité et la motivation artistique, ainsi que l'identité lorsque les ordinateurs participent au processus et travaillent d'une manière qui pourrait être définie comme créative - du moins selon certaines définitions.

⁷⁴⁹ Cibrián E. *et al.* « Artificial intelligence and machine learning for commercial analysis in the audiovisual sector: A case study of designing TV schedules », <http://www.kr.inf.uc3m.es/artificial-intelligence-and-machine-learning-for-commercial-analysis-in-the-audiovisual-sector-a-case-study-of-designing-tv-schedules/>.

8.3. L'IA est-elle différente des technologies précédentes ?

Comment empêcher l'IA de causer des préjudices ? Comment atténuer les risques liés à la création de tels systèmes d'IA ? Le monde est-il sûr si des voitures autonomes circulent dans nos rues ? Ou, le monde est-il juste si les crimes sont détectés en fonction de l'endroit où vous vivez ou de la façon dont vous regardez la caméra ?

Avant de nous lancer dans la réglementation de l'IA, nous devrions nous demander en quoi elle est différente. Qu'est-ce qui rend l'IA si spéciale qu'elle nécessiterait une attention particulière de la part des instances de régulation ? Les problèmes sont-ils vraiment nouveaux, envisagés sous l'angle de l'homme et de la société ? Ou s'agit-il simplement de nouvelles versions des mêmes problèmes fondamentaux que la loi peut déjà traiter ?

8.3.1. Qui est responsable lorsque l'IA cause un préjudice ?

Le concept clé découlant de la question de savoir qui est responsable lorsque l'IA cause un préjudice est la capacité supposée de l'IA à agir comme un être humain sans la responsabilité ni le contrôle associés aux actes humains. En d'autres termes : si l'IA a une autonomie potentielle, doit-elle être contrôlée et réglementée ?

Une autre particularité est le fait que les actions de l'IA sont difficiles à prévoir lors de la conception des systèmes. C'est particulièrement vrai si les systèmes sont censés être « créatifs », comme on peut s'y attendre dans le secteur de l'audiovisuel. Que se passe-t-il si les systèmes produisent des résultats si inattendus qu'il est impossible de dire que le concepteur du système aurait dû les prévoir et qu'il en est donc responsable ?

Généralement et traditionnellement, nous ne considérons pas les machines comme étant responsables de leurs actes. La partie responsable est celle qui utilise la machine ou celle qui l'a construite - sans faire suffisamment attention⁷⁵⁰.

En théorie, il est raisonnable d'imaginer un monde où les machines auraient des responsabilités et devraient réparer les préjudices qu'elles causent. Nous avons des structures comme les entreprises dotées d'une personnalité juridique qui peuvent avoir des droits et des devoirs.

Cependant, dans la plupart des cas, nous considérons que les humains sont les mieux placés pour assumer les responsabilités morales et pondérer les valeurs éthiques. Dans toute technologie haut de gamme, il est cependant difficile de localiser l'acteur pertinent responsable de résultats qui peuvent ne se produire que plus tard et dans des contextes auxquels le concepteur initial n'a jamais pensé.

Comme l'indique le Groupe européen d'éthique des sciences et des nouvelles technologies (GEE) dans son rapport « L'intelligence artificielle, la robotique et les systèmes "autonomes" », l'autonomie est un aspect important de la dignité humaine qui ne devrait

⁷⁵⁰ Petit N., *op.cit.*

pas être relativisé et transféré à la technologie. Les machines ne peuvent pas assumer les positions morales des humains ni hériter de la dignité humaine. Toutes nos institutions morales et juridiques sont fondées sur l'idée que l'être humain assume des responsabilités morales : il doit rendre des comptes, s'acquitter de ses devoirs ou encore peut être tenu pour responsable⁷⁵¹.

8.3.2. Ce n'est pas qu'une question d'économie

Il est incontestable que la personnalisation est utile aux individus, mais dans une industrie ayant une grande pertinence culturelle et sociétale, il est crucial que ces systèmes servent des intérêts qui dépassent ceux de l'individu, en d'autres termes les intérêts du public ou de la société dans son ensemble. Il est essentiel de trouver un équilibre acceptable entre ces intérêts potentiellement concurrents. Ces derniers temps, certains des plus grands défis que les démocraties ont dû relever ont été créés par des « machines à cibler » placées entre de mauvaises mains : les « fausses nouvelles » à grande échelle, les *deepfakes* (faux parfaits), ou la capacité de créer une bulle de filtres dans laquelle vivre.

Les industries audiovisuelles revêtent une importance particulière dans les processus démocratiques de nos sociétés, tant au niveau national qu'europpéen. Et nous vivons sur une scène mondiale, en particulier pour ce qui est de l'IA. Les services de données et de logiciels sont de plus en plus fréquemment proposés par le biais de systèmes mondiaux et souvent sur un marché contrôlé par des entreprises américaines ou chinoises. La façon dont un secteur créatif et audiovisuel durable peut être encouragé, alors que l'utilisation stratégique de l'IA sur le terrain gagne en importance, est également une question de compétitivité européenne et définira le type de rôle de « puissance douce » que l'Europe aura à l'avenir.

Le secteur de l'audiovisuel dans son ensemble est étroitement lié aux besoins culturels et politiques des sociétés. Cependant, si la culture a des valeurs sociétales et une pertinence plus élevées, elle a également une valeur économique sur le marché et dans la réalité économique. Certains divisent l'impact économique de la culture entre une « valeur institutionnelle » liée aux effets macroéconomiques sur les économies au niveau national ou international, et une valeur « microéconomique », qui est la base des transactions uniques par lesquelles les gens paient pour des expériences culturelles. La valeur institutionnelle est la base théorique sur laquelle les subventions économiques sont accordées aux industries⁷⁵² culturelles.

Le concept d'IA pour le bien social (AI4SG) peut être source d'inspiration pour la conception de systèmes qui favorisent la valeur sociétale. Ces derniers incluent la confiance

⁷⁵¹ Groupe européen d'éthique des sciences et des nouvelles technologies, « Déclaration sur L'intelligence artificielle, la robotique et les systèmes "autonomes" », Commission européenne, Direction générale de la recherche et de l'innovation, http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018_fr.pdf.

⁷⁵² La Torre M., « Defining the audiovisual industry », in La Torre M. (éd.) *The economics of the audiovisual industry: Financing TV, film and web*. Palgrave Macmillan UK, pp. 16–34.

comme base de la société et l'utilisation de toutes les technologies dans le but d'avoir un impact positif sur la vie⁷⁵³.

Parmi les autres recommandations, citons le développement par étapes des systèmes et leur test en laboratoire avant de les lancer en production, ainsi que la prévention des manipulations par des interventions extérieures. La promotion du bien social est souvent réussie lorsque les utilisateurs sont impliqués dans la conception des systèmes, que l'autonomie de l'utilisateur est respectée et sert de base à la conception du système, et que l'utilisateur a la possibilité de donner un sens et une signification aux systèmes et à leurs résultats.

8.4. Nous avons l'obligation morale de faire le bien avec l'IA

Lorsque nous considérons la meilleure façon de réglementer une industrie, il convient d'arriver à un compromis entre les coûts de la réglementation et ses avantages. Si nous ne nous focalisons que sur les risques, et essayons d'éviter les préjudices et d'établir la responsabilité par le biais de la réglementation, nous pouvons perdre certains des avantages, car la réglementation décourage généralement l'innovation et entraîne des coûts⁷⁵⁴.

Le problème d'une réglementation stricte est qu'elle limite généralement la prise de risque, la créativité et la création de valeur. Nous devons donc nous demander quel niveau de risque est acceptable et qui doit assumer le risque si quelque chose tourne mal.

Lors de l'élaboration de la réglementation, il est donc essentiel d'examiner les deux aspects du phénomène. Sur le plan philosophique, il faut se demander comment inspirer l'innovation et la créativité, et comment ne pas établir un système qui décourage la prise de risques. Le risque est un carburant pour l'économie et le bien-être. La prise de risque fait également partie de la créativité et de la création de valeur dans les industries audiovisuelles.

Par conséquent, tant dans le domaine de l'IA en général que dans celui des industries audiovisuelles, avant de réglementer un ensemble spécifique - ou dans ce cas, non spécifique - de technologies, il faut s'interroger sur la réalité du problème. Et en allant encore plus loin : une réglementation devrait-elle être supprimée pour permettre de faire le bien avec l'IA ? Ou une réglementation devrait-elle être adoptée pour permettre un accès plus facile aux données ou aux droits nécessaires au passage vers une nouvelle ère industrielle, y compris dans les industries audiovisuelles ?

La discussion sur la réglementation de l'IA trouve son origine dans le domaine des instincts et des sentiments humains fondamentaux qui nous font craindre que les machines puissent prendre le contrôle ultime sur nous. C'est notamment le cas de l'intelligence

⁷⁵³ Taddeo M., « Trusting digital technologies correctly », *Minds and Machines*, 27(4), 565–568. Taddeo M., Floridi, L., « The case for e-trust », *Ethics and Information Technology*, 13(1), 1-3.

⁷⁵⁴ Gurkaynak G., Yilmaz I. et Haksever G., « Stifling artificial intelligence: Human perils », *Computer Law & Security Review*, 32(5), pp. 749–758.

artificielle générale imaginée avec une volonté propre. Cette capacité à agir indépendamment des humains et la construction en partie fictive d'une intelligence de type humain sont à la base des prédictions d'apocalypse et sous-tendent les appels à une réglementation spécifique de l'IA.

Si la réglementation est souvent considérée comme nécessaire pour gérer les risques et prévenir les préjudices, dans le cas des industries audiovisuelles, elle pourrait également être envisagée sous un autre angle éthique : nous avons l'obligation morale et éthique de faire le bien ; comment pouvons-nous encourager le bon usage de l'IA de manière responsable, et comment la réglementation pourrait-elle créer un environnement qui favorise la créativité et l'innovation, en somme : une vie européenne plus riche et la prospérité économique ?.

La réglementation est généralement établie lorsqu'il y a un problème à résoudre. Mais nous ne pouvons perdre quelque chose que si nous l'avons d'abord gagné.

8.5. La réglementation doit être centrée sur l'homme et fondée sur des objectifs

N'oublions pas que dans un monde où les risques abondent, certains d'entre eux sont souhaitables : la prise de risque est à la base de la créativité et du bien-être. Fondamentalement, la question de la réglementation de l'IA dans les industries audiovisuelles revient à savoir quels risques sont si excessifs qu'ils méritent d'être réglementés⁷⁵⁵. Si nous voulons qu'un système d'IA soit créatif, devons-nous lui permettre de créer des résultats imprévus ? Quelles sont les conséquences réelles de ces risques s'ils sont pris⁷⁵⁶ ?

L'application mécanique de règles éthiques peut parfois s'avérer délicate. Par exemple, exiger la transparence peut déboucher sur une situation absurde. En ce qui concerne le travail créatif dans le secteur de l'audiovisuel, qui a jamais su comment un artiste est arrivé à une conclusion artistique particulière ? Et voulons-nous seulement le savoir ? Le mystère ne fait-il pas partie de la fascination ? Pourquoi exiger la transparence de la part des machines utilisées dans la partie créative de l'industrie si nous n'exigeons pas la même chose des humains ? Dans ce cas, il n'y a aucun intérêt public ou privé à protéger quelqu'un de la créativité à l'aide de la technologie.

Par ailleurs, les avantages de l'utilisation de l'IA dans les industries audiovisuelles peuvent être renforcés par une réglementation efficace, car elle minimise les risques découlant d'un manque de clarté. Existe-t-il des obstacles dans le système réglementaire actuel qui devraient être supprimés ? Devrions-nous encourager des accords réglementés

⁷⁵⁵ Buiten M. C., « Towards Intelligent Regulation of Artificial Intelligence ». *European Journal of Risk Regulation*. 10 (1): 41–59, <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/towards-intelligent-regulation-of-artificial-intelligence/AF1AD1940B70DB88D2B24202EE933F1B>.

⁷⁵⁶ Scherer M. U., « Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies ». *Harvard Journal of Law & Technology*, Vol. 29, No. 2, Spring 2016, p. 364, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777.

de partage de données afin de faciliter la création de nouvelles entreprises et de générer plus de valeur pour les clients et les citoyens ?

À mon avis, l'Europe souffre d'une approche de la réglementation trop axée sur le risque. Alors que nous en sommes encore à évaluer les risques et à envisager de les gérer, les acteurs américains et chinois sont déjà entrés sur le terrain, et mettent en place des défenses pour protéger leurs rôles déjà dominants dans l'industrie.

8.5.1. Les risques majeurs devraient être pris en compte

Notre technologie nous permet déjà, ensemble, de détruire la nature et même l'humanité entière. Selon Huber, il s'agit de « risques publics » qui sont si largement dispersés et échappent tellement à la compréhension et au contrôle du porteur de risque individuel qu'ils constituent des menaces pour la santé et la sécurité humaines⁷⁵⁷. Dans le secteur de l'audiovisuel, les risques susceptibles de causer le plus de préjudices sont en fait de ce type, et ils entraînent la destruction involontaire des fondements de nos sociétés parce que les citoyens sont manipulés et leurs comportements sont orientés à des fins commerciales et politiques. D'une certaine manière, c'est un exemple de la « tragédie des biens communs »⁷⁵⁸. Selon ce concept qui est toujours débattu, les individus agissant selon leur propre intérêt se comportent contrairement au bien commun.

Une approche de cet ensemble de risques consiste à utiliser les droits de l'homme comme un objectif et un outil politique. Notre système juridique impose l'obligation d'éviter toute violation des droits de l'homme. Cependant, si la démocratie est en danger à cause d'un usage accru du ciblage, qui est responsable ? Et même si le concept de responsabilité pouvait être créé, les juges et les systèmes juridiques nationaux sont-ils capables d'identifier les parties responsables suffisamment vite pour assurer l'efficacité de la réglementation ?

Les systèmes d'apprentissage automatique ont été accusés à plusieurs reprises de prendre des décisions ou de faire des prédictions biaisées – et il a été établi que c'était effectivement le cas⁷⁵⁹. Il est en fait possible que le système soit biaisé parce qu'il a été conçu pour l'être. Le plus souvent, cependant, le biais provient des données utilisées dans la formation du système. Et enfin, ces données ne font souvent que refléter et révéler les biais de la réalité actuelle et des décisions prises par les humains. Elles reproduisent donc nos valeurs et nos choix sociétaux actuels, ainsi que les biais humains et culturels. Ceci, une fois de plus, reflète la nature sociétale des systèmes d'IA, en particulier dans les industries audiovisuelles. Nous devons faire attention à ne pas blâmer les systèmes d'IA

⁷⁵⁷ Huber P., « Safety and the second best: The hazards of public risk management in the courts », *Columbia Law Review*, 85(2), pp. 277–337.

⁷⁵⁸ Feeny D. *et al.*, « The tragedy of the commons: Twenty-two years later », *Human ecology*, 18(1), pp. 1–19.
Hardin G., « The tragedy of the commons », *Journal of Natural Resources Policy Research*. Taylor & Francis, 1(3), pp. 243–253.

Stavins R. N., « The problem of the commons: Still unsettled after 100 years », *The American economic review*, 101(1), pp. 81–108.

⁷⁵⁹ Voir les chapitres 1, 2 et 3 de cette publication.

pour quelque chose qui pourrait en fait avoir une valeur sociétale : révéler à quel point les décisions prises par les humains jusqu'à présent sont biaisées. Nous devrions plutôt saluer les systèmes, car ils nous font davantage prendre conscience de nos valeurs – en veillant, bien sûr, à ce que des décisions non éthiques ne soient pas appliquées dans la pratique avant que les tests ne soient effectués et les résultats analysés.

8.5.2. Les humains sont les responsables

Nous ne pouvons éviter de conclure qu'un système réglementaire qui fonctionne bien dans un monde en mutation rapide doit être souple et reposer sur des principes fondamentaux plutôt que constituer une tentative de régulation de la technologie. Il existe déjà des approches qui pourraient bien fonctionner : mettre l'accent sur les objectifs et se centrer sur l'être humain et non sur la technologie.

L'un des principes fondamentaux sur lesquels s'appuyer est la question de l'attribution de la responsabilité juridique. Dans toutes les sociétés, il existe un système dans lequel la responsabilité morale, la responsabilité juridique, voire dans certains cas l'obligation de rendre des comptes, est attribuée dans les cas où des préjudices sont causés à autrui. Dans nos systèmes actuels, cette responsabilité ne peut pas être attribuée à des technologies ou à des machines, mais à des personnes, des entreprises et à d'autres entités juridiques. Dans certaines industries, les risques ont été jugés si énormes que les organisations assument une responsabilité plus lourde en matière de résultats, même en l'absence de négligence. L'IA est créée par des personnes cultivées et souvent conscientes des risques possibles que leur technologie peut engendrer pour les utilisateurs ou les objets. Il n'est pas inconcevable pour elles – ou en pratique pour les institutions qu'elles représentent – d'assumer la responsabilité de leurs actes, voire dans certains domaines d'assumer une responsabilité stricte qui s'applique même sans faute.

Jusqu'à présent, chaque fois que la réglementation semble porter sur la technologie, elle concerne en réalité les personnes qui ont créé ou utilisé cette technologie, et les organisations dans lesquelles elles sont employées. Et, fait intéressant, si la technologie basée sur l'IA peut avoir ses propres défauts, les humains sont souvent encore plus défaillants. Le problème est que nous nous sommes habitués aux défauts humains mais que nous considérons que les mêmes défauts causés par la technologie sont moins acceptables⁷⁶⁰. Une question essentielle est de savoir s'il existe réellement un nouveau risque créé par l'IA, peut-être même encore inconnu⁷⁶¹.

⁷⁶⁰ Petit N., *op.cit.*

⁷⁶¹ Scherer M. U., *op.cit.*, p. 364.

8.5.3. La transparence comme solution provisoire ?

Sans transparence, les citoyens et les consommateurs sont confrontés à des décisions qu'ils ne comprennent pas et sur lesquelles ils n'ont aucun contrôle. Pour évaluer si la responsabilité d'une décision fondée sur l'IA doit être engagée, les tribunaux doivent également comprendre comment l'IA a pris sa décision. Ainsi, exiger la transparence et l'explicabilité pourrait être une solution appropriée à certains des problèmes éthiques et juridiques, et a déjà été recommandé comme outil de régulation⁷⁶². En fait, dans le débat portant sur une réglementation durable de l'IA, la transparence est un concept central, au moins comme solution provisoire.

La transparence en matière de données renvoie à l'obligation de tenir les utilisateurs, les clients ou les consommateurs informés de la façon dont leurs données sont utilisées. Dans le monde des algorithmes, la transparence signifie la nécessité d'expliquer leur mode de fonctionnement de sorte qu'ils soient compréhensibles pour les utilisateurs. L'explicabilité – un concept proche de la transparence – signifie même, en définitive, que les valeurs sur lesquelles les systèmes fondent leur comportement devraient être traçables.

Toutefois, un équilibre doit être trouvé entre les avantages et les coûts de cette transparence. Parfois, l'exigence de transparence peut, dans la pratique, rendre la mise en œuvre du système techniquement irréalisable⁷⁶³. Exiger plus de transparence pourrait même nous obliger à accepter que les systèmes soient moins précis qu'ils ne pourraient l'être techniquement. Nous nous trouvons face à un choix intéressant : si un système de diagnostic est plus précis sous la forme d'une boîte noire inexplicable, préféreriez-vous sauver une vie grâce à ce système – sans savoir comment ? Une réglementation binaire, exigeant la transparence dans tous les cas où l'IA est appliquée, n'est pas non plus envisageable.

La transparence pose également d'autres problèmes. Certains secrets commerciaux peuvent ne pas pouvoir être révélés ou les coûts du maintien de la transparence peuvent entraîner une concentration des industries qui n'est pas souhaitable⁷⁶⁴. Exiger une transparence coûteuse peut avoir des effets négatifs sur l'innovation⁷⁶⁵. Parfois, la logique des machines ne peut tout simplement pas être exprimée dans un langage compréhensible par les humains. Et comment peut-on déterminer la responsabilité dans un domaine technologique largement basé sur le partage à l'échelle mondiale des ressources telles que les morceaux de code et les modèles d'algorithme d'IA ?⁷⁶⁶

⁷⁶² Voir le chapitre 1 de cette publication.

⁷⁶³ Buiten M. C., *op.cit.*

⁷⁶⁴ Scherer M. U., *op.cit.*

⁷⁶⁵ Buiten M. C., *op.cit.*

⁷⁶⁶ Leonelli S., « Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems », *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 374(2083), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124067/>.

8.6. Centré sur l'humain, et non sur la technologie

Le débat sur la réglementation de l'IA a suivi deux voies principales : l'une basée sur le point de vue du système juridique ; l'autre alignée sur la notion de partir des technologies pour construire les besoins réglementaires de bas en haut à partir d'applications spécifiques de l'IA⁷⁶⁷. Cependant, la base d'un système réglementaire efficace et fonctionnel ne repose pas sur ces concepts mais sur les besoins des hommes et des sociétés.

Si toutes ces discussions sur les caractéristiques particulières de la technologie de l'IA sont précieuses, l'IA n'est jamais qu'une technologie – un logiciel informatique dont le domaine principal est le calcul mathématique – et il est juste de se demander si elle est aussi fondamentalement novatrice que ce qui est souvent présenté⁷⁶⁸.

En conclusion, se centrer sur l'humain en matière de réglementation de l'IA peut signifier :

- que la transparence sert les besoins humains pour donner un sens au fonctionnement des systèmes et confier les responsabilités aux bonnes personnes ;
- accepter que des risques inconnus puissent être impossibles à réglementer, du moins si la réglementation est basée sur la technologie et non sur des objectifs ;
- que nous devrions essayer de réduire le risque public sans détruire la créativité et l'innovation ;
- que nous devrions examiner l'environnement juridique actuel, en particulier en Europe, et supprimer les obstacles inutiles à l'utilisation des données pour faire le bien et créer du bien-être ;
- que les êtres humains devraient avoir un meilleur contrôle sur la façon dont leurs données sont utilisées ; cependant, le cadre actuel du RGPD ne fonctionne pas bien dans ce sens⁷⁶⁹, car les consentements sont en réalité donnés sans réelle compréhension et une grande partie de l'utilisation des données à caractère personnel se fait par l'intermédiaire de tiers ; l'Union européenne devrait revoir de manière critique ses politiques dans le domaine de l'IA et des données, et se concentrer davantage sur la possibilité de faire le bien, sans oublier de traiter les risques majeurs, en particulier pour les démocraties.

⁷⁶⁷ Petit N., *op.cit.*

⁷⁶⁸ Edelman R. D., « Here's how to regulate artificial intelligence properly », <https://www.post-gazette.com/opinion/Op-Ed/2020/01/14/R-David-Edelman-Here-s-how-to-regulate-artificial-intelligence-properly/stories/202001140013>.

⁷⁶⁹ Voir le chapitre 2 de cette publication.

Remarques finales

L'expression de bon sens selon laquelle « un grand pouvoir implique de grandes responsabilités » (rendue célèbre par la bande dessinée « Spider-Man » mais qui remonte au moins à la Révolution française⁷⁷⁰) va comme un gant à l'IA. L'IA a un potentiel énorme, qui peut être mis au service du bien comme du mal. C'est pourquoi elle fascine et effraie à la fois, et, alors que certains auront tendance à s'inquiéter, d'autres mettront l'accent sur toutes les choses merveilleuses que cette technologie révolutionnaire pourrait permettre de réaliser. En effet, la lecture des différentes contributions rassemblées dans ce rapport montre qu'il n'existe pas un point de vue unique sur la façon dont l'IA devrait être réglementée, et pourtant certains principes semblent être (d'une façon ou d'une autre) présents à l'esprit de tous les auteurs : l'explicabilité, la confiance, le respect de la vie privée, le pluralisme, mais aussi la liberté d'expression, la créativité et l'innovation. Si tous ces objectifs pouvaient être combinés, l'IA pourrait représenter à plusieurs égards une bénédiction pour l'humanité.

À moins bien sûr qu'un jour, les pires cauchemars d'Elon Musk ne deviennent réalité et que les machines prennent le contrôle du monde. Mais un tel avenir dystopique n'est pas à l'horizon.

Tout du moins pas encore.

⁷⁷⁰ <https://quoteinvestigator.com/2015/07/23/great-power/>

Une publication
de l'Observatoire européen de l'audiovisuel

