

IA et secteur audiovisuel : exploration du paysage juridique actuel

IRIS

Une publication
de l'Observatoire européen de l'audiovisuel

IRIS**IA et secteur audiovisuel : exploration du paysage juridique actuel**

Observatoire européen de l'audiovisuel, Strasbourg, 2024
ISSN 2079-1070

Directrice de publication – Susanne Nikoltchev, Directrice exécutive

Supervision éditoriale – Maja Cappello, Responsable du Département Informations juridiques

Équipe éditoriale - Justine Radel-Cormann, Sophie Valais

Observatoire européen de l'audiovisuel

Auteurs (en ordre alphabétique)

Malte Baumann, Judit Bayer, Mira Burri, Gianluca Campus, Mark Cole, Kelsey Farish, Philipp Hacker, Elodie Migliore, Jan Bernd Nordemann, Justine Radel-Corman, Bart van der Sloot

Relecture

Aurélie Courtinat, Udo Lücke, Anthony Mills

Traduction

Julie Mamou, Nathalie Sturlèse, Erwin Rohwer

Assistante éditoriale – Sabine Bouajaja

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

Observatoire européen de l'audiovisuel

L

Éditeur

Observatoire européen de l'audiovisuel
76, allée de la Robertsau, 67000 Strasbourg, France
Tel.: +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.), IA et secteur audiovisuel : exploration du paysage juridique actuel , IRIS, Observatory, Strasbourg, octobre 2024

© Observatoire européen de l'audiovisuel (Conseil de l'Europe), Strasbourg, 2024

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, de ses membres ou du Conseil de l'Europe.

IA et secteur audiovisuel : exploration du paysage juridique actuel

**Malte Baumann, Judit Bayer, Mira Burri, Gianluca Campus, Mark Cole, Kelsey Farish,
Philipp Hacker, Elodie Migliore, Jan Bernd Nordemann, Justine Radel-Corman, Bart
van der Sloot**

Avant-propos

A la question de savoir si la création assistée par IA était en train de devenir un nouveau genre cinématographique, les créateurs Pierre Zandrowicz et Matt Tierney ont répondu « *cela nous donne simplement davantage de pinceaux dans notre pot, ce qui veut dire que ce que nous pouvons faire concrètement, c'est prendre chaque plan du film et le peindre à partir de texte, grâce à des prompts* »¹.

Cette déclaration illustre le rôle de l'IA, en particulier de l'IA générative, en tant qu'outil permettant à divers professionnels d'améliorer leur travail. L'IA générative peut permettre aux travailleurs de devenir plus multidisciplinaires ; par exemple, les auteurs peuvent utiliser un outil d'IA générative pour créer des visuels afin de promouvoir leur travail. Toutefois, l'essor de l'IA générative dans le secteur audiovisuel s'accompagne aussi de nouveaux enjeux, tels que des risques pour l'emploi et les droits d'auteur, auxquels les décideurs doivent répondre.

Dans ce contexte, l'Observatoire européen de l'audiovisuel (OEA) a rouvert son dossier de 2020 sur l'IA afin d'explorer l'intersection entre innovation technologique et cadres législatifs. Ce rapport aborde certaines des difficultés que les réglementations existantes rencontrent en matière d'IA dans le secteur audiovisuel, et cherche à déterminer si ces réglementations seront en mesure de faire face à l'épreuve du temps et de s'adapter à l'évolution du paysage technologique. Conçu, élaboré et coordonné par le département juridique de l'OEA, ce rapport est divisé en quatre parties.

La première partie familiarise les lecteurs avec la question de l'IA dans le secteur audiovisuel, en présentant à la fois ses avantages et les complications auxquelles le cadre réglementaire (plutôt) fragmenté devra remédier. Le **chapitre 1**, rédigé par Justine Radel-Cormann (OEA), pose les bases de cette discussion.

La deuxième partie se penche sur les questions juridiques entourant l'IA et les données qui l'alimentent. Le **chapitre 2**, de Philipp Hacker (Université de Yale), explore ses implications en matière de protection des données et de la vie privée, ainsi que l'impact de règlements tels que le RGPD et le Règlement sur l'IA. Il examine également la question des transferts internationaux de données et présente un comparatif avec le droit américain. Le **chapitre 3**, rédigé par Gianluca Campus (PwC Digital Innovation), analyse les sujets de l'utilisation d'œuvres protégées par le droit d'auteur pour l'apprentissage de l'IA, la création d'œuvres dérivées et le cadre juridique applicable à l'utilisation de données protégées par le droit d'auteur.

La troisième partie aborde cinq questions clés que l'IA pose au secteur audiovisuel. Le **chapitre 4**, rédigé par Malte Baumann et Jan Nordemann (cabinet d'avocats NORDEMANN, Berlin), traite des questions de paternité, de responsabilité et de

¹ [Interview](#) des créateurs Pierre Zandrowicz et Matt Tierney par Alex Helisek de Breezeway Productions lors du festival du film de Tribeca de 2023. [Leur film d'animation « In Search of Time » avait été présenté au festival de Tribeca en 2023.](#)

transparence à l'ère de l'IA générative. Le **chapitre 5**, rédigé par Kelsey Farish (Reviewed & Cleared, Londres), traite de la protection de l'image, de la voix et des droits de la personnalité des acteurs vis-à-vis de la reproduction par IA. Le **chapitre 6**, rédigé par Elodie Migliore (Université de Strasbourg), examine l'impact de l'IA sur le droit du travail, en se référant aux récentes grèves et initiatives législatives observées aux Etats-Unis. Le **chapitre 7**, rédigé par Judit Bayer (Université de Münster), étudie le rôle de l'IA dans la désinformation et les mesures réglementaires pour la combattre. Le **chapitre 8**, rédigé par Mira Burri (Université de Lucerne), explore l'impact de l'IA sur le pluralisme des médias et la diversité culturelle (par exemple, les problèmes portant sur la personnalisation et l'impartialité des contenus) et les mesures réglementaires possibles pour atténuer ces effets et promouvoir une consommation diversifiée des contenus.

La quatrième partie se tourne vers l'avenir et tente de déterminer si les réglementations récentes en matière d'IA sont en mesure de relever les défis posés par l'IA au secteur audiovisuel. Le **chapitre 9**, rédigé par Mark Cole et Sandra Schmitz (EMR), offre une perspective sur la manière dont de futures réglementations pourraient mieux répondre aux difficultés et aux opportunités amenées par l'IA, en garantissant une approche équilibrée qui encourage l'innovation tout en protégeant les droits et les intérêts de toutes les parties prenantes de l'industrie audiovisuelle. Le **chapitre 10**, rédigé par Bart van der Sloot (Université de Tilburg), conclut la publication en abordant des dilemmes éthiques tels que l'authenticité, le risque de distorsion de la réalité par l'IA et les impacts sociétaux plus généraux des contenus générés par l'IA.

Les remarques d'introduction et de conclusion, rédigées par Justine Radel-Cormann (OEA), visent à contextualiser ces diverses questions juridiques et politiques.

Je remercie chaleureusement les brillants auteurs qui ont contribué à ce riche rapport. A nos lecteurs, je souhaite tout simplement une bonne lecture !

Strasbourg, octobre 2024

Maja Cappello

Coordinatrice IRIS

Responsable du Département Informations juridiques

Observatoire européen de l'audiovisuel

Table des matières

1. L'intelligence dans le secteur audiovisuel.....	2
1.1. La définition de l'« IA » et d'« audiovisuel »	4
1.2. Les possibilités transformatives de l'IA dans l'industrie audiovisuelle	7
1.3. Exemples d'utilisation de l'IA dans l'industrie audiovisuelle	8
1.3.1. Étude de cas 1 : Claude, un assistant conversationnel pour le développement de projets.....	9
1.3.2. Étude de cas 2 : DiversityCatch, pour mesurer la diversité du contenu	10
1.3.3. Étude de cas 3 : Midjourney et DALL·E, outils d'IA pour la création d'images et de vidéos	10
1.4. Défis posés à l'industrie audiovisuelle par l'IA	11
1.4.1. Quels sont les défis à venir pour le secteur audiovisuel ?	11
1.4.2. Le cadre législatif de l'IA : un puzzle complexe.....	13
<hr/>	
2. L'IA et la protection des données dans les médias audiovisuels.....	18
2.1. Introduction	18
2.2. Les contenus audiovisuels en tant que données à caractère personnel	19
2.3. Les problématiques en matière de protection des données et de confidentialité	21
2.3.1. La base juridique de l'entraînement.....	22
2.3.2. Les hallucinations.....	23
2.3.3. Les LLM en tant que données à caractère personnel.....	24
2.3.4. Les données sensibles	25
2.3.5. Information et capacité de contrôle des utilisateurs et utilisatrices.....	26
2.3.6. La prise de décision automatisée	27
2.4. Le règlement sur l'IA.....	28
2.5. Les transferts internationaux de données	29
2.6. Comparaison avec le droit américain et le droit international	30
<hr/>	
3. Entraînement de l'IA et préservation des droits d'auteur	33
3.1. Introduction	33
3.1.1. Aperçu des systèmes d'IA et de leur traitement des données protégées par le droit d'auteur	33
3.1.2. Considérations sur les œuvres dérivées.....	35
3.2. Exception pour la fouille de textes et de données pour les données d'apprentissage	36
3.2.1. Applicabilité de l'exemption au titre de la fouille de textes et de données pour les données d'apprentissage de l'IA	36
3.2.2. La fouille de textes et de données et l'impact sur les droits de reproduction et d'extraction.....	38

3.3. Textes législatifs relatifs à l'IA	40
3.3.1. Règlement de l'UE sur l'IA et droit d'auteur : règles de transparence et mesures relatives à la fouille de textes et de données	40
3.3.2. IA et exception relative à la fouille de textes et de données : les propositions de lois de l'Italie et de la Pologne	41
3.4. L'impact de la jurisprudence.....	42
3.4.1. Vue d'ensemble des affaires relatives aux données d'apprentissage (États-Unis et Europe).....	42
3.5. Quelques observations (préliminaires) sur la jurisprudence.....	44

4. Paternité des œuvres, responsabilité et transparence dans le contexte des contenus générés par l'IA..... 47

4.1. Paternité.....	47
4.1.1. Le créateur humain en tant qu'auteur	47
4.1.2. La création d'œuvres assistée par l'IA	49
4.1.3. Qui est l'auteur ?	51
4.1.4. La protection par les droits voisins.....	51
4.2. La responsabilité sur les produits générés par l'IA	52
4.2.1. Quand faut-il considérer qu'une infraction a été commise ?	52
4.2.2. Exceptions et limitations au droit d'auteur applicables aux contenus générés par l'IA.....	54
4.2.3. Responsabilité de l'utilisateur	55
4.2.4. Les conditions d'utilisation des fournisseurs d'IA.....	58
4.2.5. Réduire la responsabilité éventuelle.....	58
4.2.6. Transparence.....	59

5. Droits de la personnalité et transparence 61

5.1. Le contexte.....	61
5.2. Les facteurs commerciaux.....	63
5.2.1. L'évolution des doubles numériques.....	63
5.2.2. Le point de vue des artistes-interprètes : autonomisation ou exploitation ?	64
5.2.3. Le vide réglementaire	65
5.3. La transparence dans les instruments européens	66
5.3.1. Le règlement européen sur l'IA	66
5.3.2. La transparence dans la Convention-cadre sur l'IA.....	69
5.4. D'autres approches : les États-Unis et le Royaume-Uni	70
5.5. La transparence, pilier du respect des droits de la personnalité	73

6. Impact de l'IA sur le marché du travail audiovisuel en Europe 75

6.1. Introduction	75
6.2. Impact de l'IA sur le droit du travail dans le secteur audiovisuel aux États-Unis	76
6.2.1. Les grèves WGA et SAG-AFTRA.....	76
6.2.2. L'accord WGA après la grève	77
6.2.3. L'accord SAG-AFTRA	78
6.3. Impacts de l'IA sur le droit du travail dans le secteur audiovisuel dans l'UE.....	80
6.3.1. L'approche de l'Union européenne	80
6.4. Analyse d'un éventail d'initiatives émanant de différents acteurs.....	82
6.4.1. Les organismes de gestion collective (OGC).....	82
6.4.2. Associations et fédérations	84
6.4.3. Organisations syndicales	85
6.5. Observations finales : les lacunes à combler et les perspectives d'avenir	87

7. Désinformation et intelligence artificielle dans le secteur audiovisuel 88

7.1. Définition de la désinformation.....	88
7.2. Les applications de l'IA dans l'industrie de la désinformation	89
7.2.1. L'IA générative.....	89
7.2.2. Textes et images.....	89
7.2.3. Hypertrucages (<i>deep fakes</i>).....	90
7.2.4. Contenus audio.....	91
7.2.5. Robots.....	92
7.2.6. IA et fausses informations.....	93
7.3. La lutte contre la désinformation	94
7.3.1. Réglementation.....	94
7.3.2. Vérification des faits avec l'aide de l'IA.....	97
7.4. Conclusion	99

8. Pluralisme et diversité..... 101

8.1. Contexte général : l'IA, une technologie disruptive.....	101
8.2. Impact de l'IA sur la liberté d'expression, le pluralisme des médias et la diversité culturelle.....	103
8.2.1. Remarques préliminaires : en quoi consiste le changement ?.....	103
8.2.2. Impact sur la diffusion et la consommation des contenus.....	104
8.2.3. Impact sur la création de contenus.....	106
8.2.4. Autres aspects à prendre en considération	108
8.3. Gérer les répercussions réelles et potentielles des systèmes d'IA sur le pluralisme et la diversité.....	110
8.4. Conclusion.....	111

9. Le monde de demain : les textes juridiques sont-ils à l'épreuve de l'intelligence artificielle et des défis de l'audiovisuel ? 112

9.1. Récapitulatif : approches réglementaires existantes et à venir	112
9.1.1. En commençant par les recommandations : approches initiales de l'OCDE et de l'UNESCO	113
9.1.2. Vers un droit contraignant : les avancées au sein du Conseil de l'Europe et de l'UE	114
9.2. Les approches réglementaires à l'épreuve de la réalité : applicabilité et limites dans le cadre du secteur audiovisuel	119
9.2.1. Aspects relatifs à la protection des données.....	119
9.2.2. Aspects relatifs aux droits de propriété intellectuelle.....	120
9.2.3. Aspects relatifs aux droits de la personnalité.....	123
9.2.4. La désinformation : un défi de taille	124
9.2.5. Portée du Règlement sur l'IA : champ d'application géographique.....	126
9.3. Les perspectives d'avenir : « l'épreuve du temps » et les normes mondiales.....	127
9.3.1. Vers des normes globales et flexibles basées sur le risque dans une législation spécifique.....	127
9.3.2. Examen des aspects spécifiques au secteur des médias (audiovisuels).....	129

10. Dilemmes éthiques et défis sociétaux posés par l'IA générative 131

10.1. Introduction	131
10.2. Nos fondements éthiques	132
10.3. Impact de l'IA sur ces fondements.....	133
10.4. Dilemmes éthiques dans le secteur audiovisuel	135
10.5. Défis sociétaux posés par l'IA.....	137
10.6. Conclusion	140

11. Conclusions..... 143

Illustrations

Schéma 1 De la sémantique réglementaire aux concepts et applications clés.....	3
Schéma 2 Exemples d'applications d'IA dans la chaîne de valeur audiovisuelle	9
Schéma 3 L'IA : exemple de législations variées	15

Tableaux

Tableau 1 Définitions	4
-----------------------------	---

PARTIE I - L'intelligence artificielle générationnelle et son potentiel de transformation du secteur audiovisuel

L'intelligence artificielle générative (IA générative) est au cœur d'une nouvelle vague frénétique d'activités réglementaires. Si les discussions au sein de l'UE en vue de l'adoption du Règlement sur l'IA avaient débuté dès avril 2021, celles-ci ont pris de l'ampleur à la suite de la mise à disposition du grand public d'un logiciel ouvert d'IA générative à la fin de l'année 2022. L'IA générative peut produire de nouveaux contenus, tels que du texte, des images, du son, des vidéos, etc. sur la base de phrases (prompts) fournies par les utilisateurs dans l'outil d'IA générative. La qualité de ces prompts influe sur la qualité du résultat.

Les possibilités introduites par l'IA générative sont infinies, avec à la fois des opportunités créatives mais aussi des gains d'efficacité. Dans le secteur audiovisuel, l'IA pourrait se révéler utile à différents stades de la chaîne de valeur. Avec l'IA générative, il existe tout un univers de possibilités, dans lequel les rôles pourraient se chevaucher, en permettant aux créateurs individuels de réaliser des tâches dépassant leur champ d'action traditionnel, favorisant ainsi une approche plus multidisciplinaire. Par exemple, un auteur pourrait-il créer une trame musicale pour son scénario ? Les tâches d'un scénariste pourraient-elles rejoindre celles d'un éditeur ? Ces rôles pourraient-ils finir par se confondre ?

Ou au contraire, cette pluridisciplinarité ne serait-elle qu'un mirage, en fin de compte peu utile aux créateurs ?



1. L'intelligence dans le secteur audiovisuel

Justine Radel-Cormann, Legal Analyst, Observatoire européen de l'audiovisuel

Le secteur audiovisuel, qui se transforme en permanence pour répondre à l'évolution des besoins et des préférences du public, occupe depuis longtemps une place de premier plan dans le domaine des avancées technologiques et numériques. Depuis les premières caméras qui captaient des films muets en noir et blanc jusqu'à l'ère moderne marquée par la diffusion en continu en ultra-haute définition sur des terminaux portables, l'industrie a su tirer parti de l'innovation pour améliorer tant la création que la distribution de contenus.

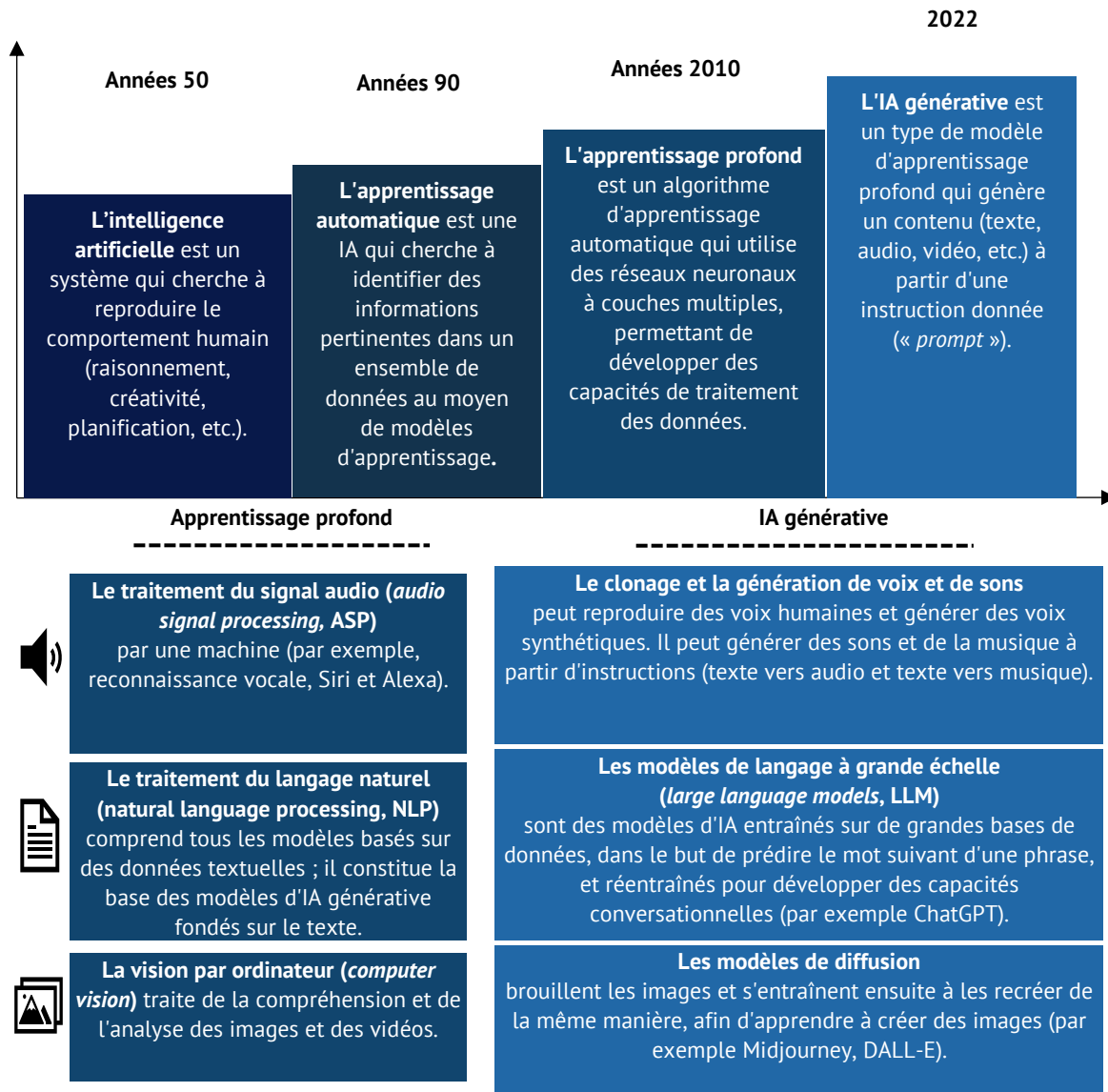
Les développements les plus récents dans le secteur concernent les nouvelles fonctions amenées par l'intelligence artificielle (IA) : l'apprentissage profond progresse depuis une décennie, et culmine avec l'essor récent de l'intelligence artificielle générative. Les applications potentielles qu'elle offre au secteur audiovisuel ont suscité à la fois enthousiasme et inquiétude.

Le schéma 1 ci-dessous illustre l'évolution de l'IA et de ses différentes technologies et applications².

² Ce visuel a été créé sur la base du [blog IBM « AI, machine learning and deep learning : what's the difference ? »](#) et du [rapport du CNC « Quel impact de l'IA sur les filières du cinéma, de l'audiovisuel et du jeu vidéo »](#) du 8 avril 2024.



Schéma 1 De la sémantique réglementaire aux concepts et applications clés



Source : Observatoire européen de l'audiovisuel

L'IA générative est en passe de remodeler l'industrie audiovisuelle, et en affecte rapidement toutes ses composantes, de la création de contenu à sa distribution. Ainsi, le paysage réglementaire doit s'adapter à son évolution rapide. La section 1 pose les définitions des termes « audiovisuel » et « IA » tels qu'ils sont utilisés dans le rapport. La section 2 examine les bénéfices que l'IA pourrait apporter à l'industrie, avec des exemples spécifiques dans la section 3. Enfin, la section 4 abordera les différents défis à venir et examinera le cadre législatif existant et ses implications.



1.1. La définition de l'« IA » et d'« audiovisuel »

Le terme « audiovisuel » se réfère par essence à tous les médias, à l'exception de la presse écrite : le cinéma, la télévision, la radio, la vidéo et les divers services à la demande (tels que la vidéo à la demande ou la télévision de rattrapage) sont tous des secteurs de l'industrie audiovisuelle. En outre, s'agissant de la chaîne de valeur, nous nous référons aux différentes branches de l'industrie audiovisuelle, telles que la production de films, les distributeurs, les exploitants et les radiodiffuseurs publics et privés³.

La notion d'« IA » est plus complexe ; il n'existe pas de consensus général sur une définition⁴. Il s'agit d'un vaste phénomène que de nombreuses parties tentent de comprendre, et il existe donc diverses définitions au niveau international (OCDE, Conseil de l'Europe), de l'Union européenne, national (États-Unis, Chine et Royaume-Uni) et de l'industrie (OpenAI, MetaAI, Gemini).

Tableau 1 Définitions

Texte	Article	Citation
Textes internationaux		
OCDE, Recommandation du Conseil⁵	Point 1	Un « système d'IA » est un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influencer sur des environnements physiques ou virtuels. Différents systèmes d'IA présentent des degrés variables d'autonomie et d'adaptabilité après déploiement.
Convention-cadre du Conseil de l'Europe⁶	Article 2	On entend par « système d'IA » un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influencer sur des environnements physiques ou virtuels. Différents systèmes d'intelligence artificielle présentent des degrés variables d'autonomie et d'adaptabilité après déploiement.
Textes de l'Union Européenne		
Règlement sur	Article 3(1)	On entend par « système d'IA », un système automatisé qui est conçu

³ Voir le considérant 23 de la Directive 2010/13/UE du 10 mars 2010 visant à la coordination de certaines dispositions législatives, réglementaires et administratives relatives à la fourniture de services de médias audiovisuels ([Directive Services de médias audiovisuels](#)) : « Aux fins de la présente directive, le terme « audiovisuel » devrait se référer aux images animées, combinées ou non à du son, et donc couvrir les films muets, mais pas la transmission audio ni les services de radiodiffusion ».

⁴ « *One of the biggest problems in regulating AI is agreeing on a definition* » (« L'un des plus grands problèmes en matière de régulation de l'IA est de parvenir à s'entendre sur une définition »), [Carnegie Endowment for International Peace, 2022](#)

⁵ [OCDE, Recommandation du Conseil sur l'intelligence artificielle](#), adoptée le 22 mai 2019, et amendée le 3 mai 2024.

⁶ [Convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit](#), adoptée le 17 mai 2024 par le Comité des Ministres du Conseil de l'Europe.



l'intelligence artificielle ⁷		pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels.
Proposition de directive sur la responsabilité en matière d'IA ⁸	Article 2(1)	On entend par « système d'IA », un système d'IA au sens du règlement sur l'IA.
Textes nationaux		
États-Unis Département d'État sur l'IA ⁹	/	Le terme « intelligence artificielle » désigne un système basé sur une machine qui peut, à partir d'un ensemble donné d'objectifs définis par l'homme, faire des prédictions, des recommandations ou prendre des décisions influençant des environnements réels ou virtuels.
Royaume-Uni Livre blanc sur la réglementation en matière d'IA ¹⁰	Point 3.2.1	Le livre blanc s'abstient de proposer une définition figée de l'IA en raison de son évolution rapide. Il se concentre plutôt sur deux caractéristiques déterminantes qui nécessitent une attention réglementaire : <ul style="list-style-type: none">• L'adaptabilité : la capacité de l'IA à s'entraîner à partir de données et à faire des déductions qui peuvent aboutir à des résultats difficiles à expliquer ou à prédire.• L'autonomie : certains systèmes d'IA peuvent prendre des décisions sans l'intention expresse ou le contrôle continu d'un être humain.
Chine Proposition de loi sur l'IA de la République populaire de Chine ¹¹	Article 94 (i)	L'IA est une technologie qui utilise des ordinateurs pour simuler le comportement humain intelligent à des fins de prédiction, de recommandation, de prise de décision ou de génération de contenu, etc. à des fins spécialisées ou générales.
Industrie		
OpenAI ¹²	Charte	Charte d'OpenAI : « La mission d'OpenAI est de s'assurer que l'intelligence artificielle générale - par laquelle nous entendons des systèmes hautement autonomes qui surpassent les humains dans la plupart des tâches économiquement utiles - bénéficie à l'ensemble de

⁷ [Règlement \(UE\) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle \(règlement sur l'IA\).](#)

⁸ [Proposition de directive sur la responsabilité en matière d'IA](#), proposée par la Commission européenne le 28 septembre 2022.

⁹ Citation du site web du [Département d'État américain sur ce qu'est l'IA](#) de 2020.

¹⁰ [Une approche pro-innovation de la réglementation de l'IA](#), présentée au Parlement par le secrétaire d'État à la science, à l'innovation et à la technologie sur ordre de Sa Majesté le 29 mars 2023.

¹¹ [Proposition de loi sur l'IA de la République populaire de Chine](#), document préliminaire qui a circulé dans les milieux universitaires, hébergé sur le site web du « *Centre for Security and Emerging Technology* » (*Georgetown University's Walsh School of Foreign Service*) et traduit en anglais.

¹² [Charte d'OpenAI](#).



		l'humanité ».
Meta AI ¹³	Page de Meta AI	Meta AI est un assistant intelligent capable de raisonner de manière complexe, de suivre des instructions, de visualiser des idées et de résoudre des problèmes complexes.
Gemini Google ¹⁴	Présentation de Gemini	Gemini a été développé dès sa conception pour être multimodal, ce qui signifie qu'il peut généraliser et comprendre, exploiter et combiner de façon fluide différents types d'informations, notamment du texte, du code, de l'audio, des images et de la vidéo.

Source : Observatoire européen de l'audiovisuel

Sur la base de ces neuf exemples, il apparaît clairement que les définitions établies par les différentes entités présentent des caractéristiques communes (terminologie, champ d'application). Néanmoins, celles-ci peuvent se concentrer sur des domaines différents en fonction du contexte et des objectifs de l'entité qui propose la définition.

Critères communs :

- Un accent sur les objectifs : les systèmes d'IA sont conçus pour atteindre des objectifs explicites ou implicites (par exemple, faire des prédictions, générer du contenu, prendre des décisions).
- L'entraînement de la machine : les systèmes d'IA sont alimentés en informations et génèrent des résultats qui peuvent influencer l'environnement physique ou virtuel.
- L'autonomie et l'adaptabilité : les systèmes d'IA varient dans leurs niveaux d'autonomie et d'adaptabilité une fois qu'ils ont été conçus ; cette variation implique que les systèmes d'IA peuvent évoluer ou apprendre à partir de leurs interactions avec des données et l'environnement.

Éléments divergents :

- La terminologie : la plupart des sources font référence aux « systèmes d'IA » ou à l'« intelligence artificielle », mais quelques-unes parlent de « systèmes basés sur des machines » ou de « technologie ».
- Ses utilisations et son influence : l'impact sur la prise de décision est mentionné une fois, certaines sources ne faisant référence qu'aux prédictions, aux recommandations et à la génération de contenu.
- La référence à l'homme : OpenAI parle d'une « IA générale » capable de surpasser les humains ; les autres sources donnent des exemples de capacités (prédictions, recommandations et génération de contenu).
- Les capacités de l'IA : les définitions vont des systèmes simulant l'intelligence humaine à ceux qui résolvent des problèmes complexes.

¹³ [Description du service Meta AI.](#)

¹⁴ [Présentation de Gemini, par Demis Hassabis, PDG et cofondateur de Google DeepMind, décembre 2023.](#)



Si la définition et les aspects techniques de l'IA peuvent être complexes pour les non scientifiques, ses applications peuvent être plus faciles à comprendre pour les non experts : les bénéfices de l'IA deviennent plus manifestes lorsqu'ils sont envisagés dans le contexte du secteur audiovisuel.

1.2. Les possibilités transformatives de l'IA dans l'industrie audiovisuelle

L'IA pourrait avoir un impact positif sur toute la chaîne de valeur de l'industrie audiovisuelle (du concept initial du contenu à sa production, sa distribution et sa protection) en contribuant au processus créatif, en automatisant des tâches, en promouvant la diversité linguistique, en améliorant la distribution du contenu, en luttant contre le piratage et en renforçant les valeurs démocratiques¹⁵.



Créativité et génération d'idées : les systèmes d'IA générative peuvent stimuler la créativité en aidant à la création et à la production de contenu. Les auteurs peuvent utiliser l'IA pour générer des idées alternatives et surmonter le syndrome de la page blanche. L'IA peut également suggérer des concepts de design et des visuels pour les plateaux de tournage et les affiches de films. Même si ces suggestions générées par l'IA ne sont pas parfaites, elles peuvent contribuer à faire avancer le processus créatif. En outre, de nombreux outils d'IA sont actuellement gratuits ou peu coûteux, ce qui permet de les rendre accessibles à un grand nombre d'utilisateurs disposant d'un accès à internet et d'un ordinateur. Cette accessibilité peut aider les créateurs disposant d'un budget limité à présenter leurs idées aux producteurs, leur permettant de lancer des processus de développement potentiels.




Automatiser les tâches administratives : l'IA peut également automatiser des tâches chronophages à faible valeur ajoutée créative, comme l'analyse des données d'audience pour analyser les préférences en matière de contenu. En outre, l'IA peut faire gagner du temps sur les tâches administratives, comme la création et la gestion des calendriers de tournage et la coordination de la logistique de l'équipe.





Sélection et personnalisation du contenu : les outils alimentés par l'IA permettent la sélection de contenus en les filtrant, les catégorisant et les classant automatiquement en fonction des centres d'intérêt du public. Cela permet de mieux cibler le contenu et d'augmenter son accessibilité en suggérant de nouveaux contenus à différents publics.


¹⁵ Pour plus d'informations sur ses différents bénéfices, voir : i) les résultats d'une enquête menée auprès des membres du Comité consultatif de l'OEA en mars 2024 ; ii) « [BBC's plans for GenAI and how we plan to use AI tools responsibly](#) », BBC, 28 février 2024 ; iii) [le rapport du CNC « Quel impact de l'IA sur les filières du cinéma, de l'audiovisuel et du jeu vidéo »](#), 8 avril 2024 ; iv) [l'enquête du DACS sur les artistes et l'IA, « AI and artists' work »](#), DACS, 18 janvier 2024 ; v) « [AI is transforming the entertainment business](#) », *The Economist*, 4 janvier 2024.



 **Traduction et diversité linguistique** : les outils de traduction assistée par l'IA peuvent accroître la diversité linguistique en rendant les contenus audiovisuels disponibles dans un plus grand nombre de langues, et en promouvoir l'accessibilité. L'utilisation d'avatars pour la traduction en langue des signes peut améliorer l'accessibilité aux téléspectateurs malentendants. Ces outils d'IA peuvent également accélérer la diffusion des contenus en les traduisant plus rapidement, ce qui leur permet d'atteindre un public plus large.

 **Lutte contre le piratage et protection du contenu** : les outils d'IA peuvent tracer l'utilisation d'œuvres protégées par le droit d'auteur, garantissant ainsi une rémunération appropriée des auteurs, et détecter toute utilisation non autorisée, ce qui permet de lutter contre les infractions. Les outils de lutte contre le piratage basés sur l'IA peuvent aider à localiser les sources de piratage et à y remédier.

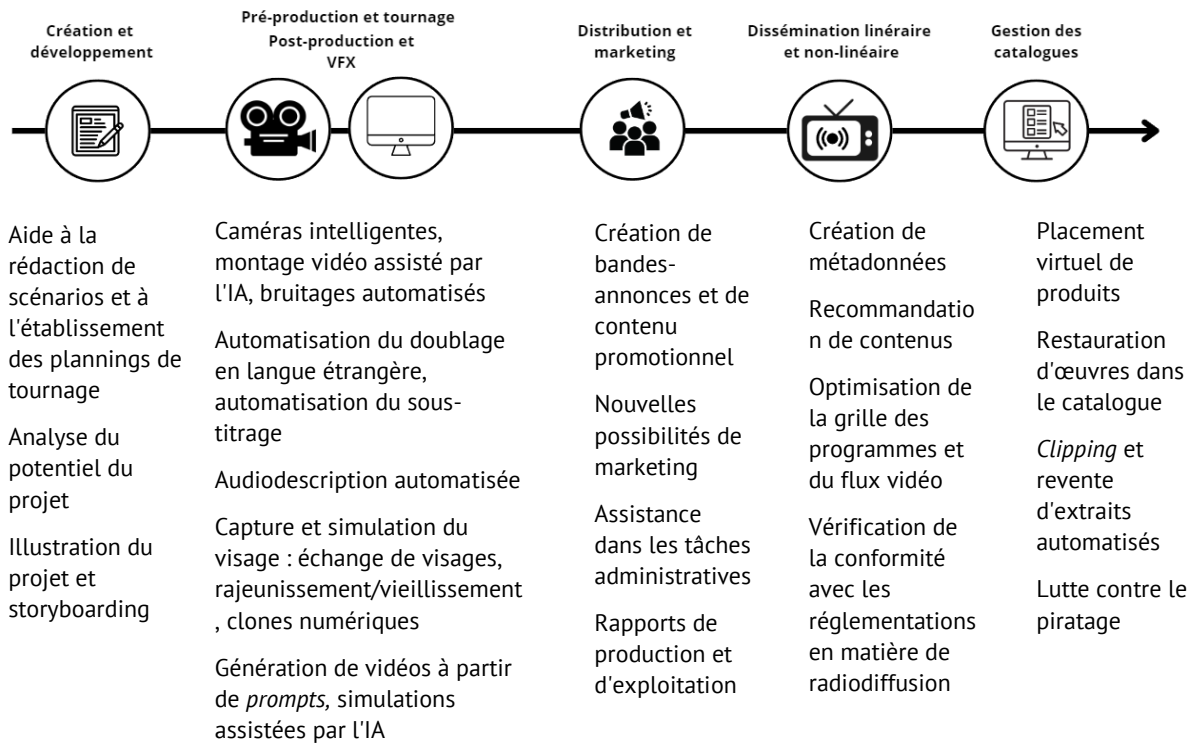
 **Promotion du pluralisme des médias** : l'IA peut contribuer à promouvoir les valeurs démocratiques en rapprochant les rédactions d'un public qui ne s'intéresse pas forcément aux médias traditionnels. Les outils d'IA peuvent donner accès à des informations fiables et diversifiées et favoriser le pluralisme des médias en offrant un contenu qui trouve un écho auprès d'un public plus large.

 **Amélioration de l'expérience du public et préservation des contenus patrimoniaux** : les outils d'IA peuvent faciliter la restauration de films anciens et améliorer leur qualité d'image en y ajoutant des pixels ou des couleurs. La restauration du son est également possible. Ces restaurations peuvent même améliorer la qualité de l'image pour des diffusions de meilleure qualité à la télévision (comme en 4K).

1.3. Exemples d'utilisation de l'IA dans l'industrie audiovisuelle

De nombreuses applications d'IA peuvent intervenir tout au long de la chaîne de valeur du contenu dans l'industrie audiovisuelle : depuis la création et le développement jusqu'à la diffusion sur des plateformes linéaires ou non linéaires. Le graphique ci-dessous fait le point sur les différentes utilisations possibles de l'IA (de façon non exhaustive)¹⁶. Il est suivi de trois exemples concrets d'outils d'IA.

¹⁶ Rapport du CNC « Quel impact de l'IA sur les filières du cinéma, de l'audiovisuel et du jeu vidéo », 8 avril 2024, et « *How genAI tools like Lore machine revisualize storyboarding* », *Variety*, 15 mars 2024.

**Schéma 2 Exemples d'applications d'IA dans la chaîne de valeur audiovisuelle**

Source : Observatoire européen de l'audiovisuel

1.3.1. Étude de cas 1 : Claude, un assistant conversationnel pour le développement de projets

Claude, un assistant conversationnel développé par Anthropic, est conçu pour faciliter le *brainstorming* et le développement d'idées¹⁷. L'utilisation de l'outil est gratuite avec certaines limitations, et des fonctionnalités supplémentaires sont disponibles moyennant un abonnement à Claude Pro. En fournissant à Claude des données telles qu'un scénario ou une histoire, les utilisateurs peuvent obtenir des idées sur divers aspects de la production du contenu, notamment sur la nécessité de procéder à des réécritures, les budgets de tournage (avec des répartitions détaillées), des suggestions de réduction des coûts, le nombre de figurants nécessaires et l'identification des scènes nécessitant une préparation spéciale ou des effets visuels. Il peut également proposer des estimations de ventes par territoire et aider à la distribution en fournissant des listes des principaux distributeurs étrangers et leurs coordonnées.

Bien que les recommandations de Claude ne soient pas toujours exactes à 100 %, elles offrent de précieuses perspectives alternatives. L'inconvénient est le manque de

¹⁷ <https://www.anthropic.com/claude>.



transparence des données, car les sources d'information qui alimentent Claude ne sont pas divulguées.

1.3.2. Étude de cas 2 : DiversityCatch, pour mesurer la diversité du contenu

Développée par MediaCatch en collaboration avec une université danoise, DiversityCatch est une solution logicielle pilotée par IA conçue pour mesurer la diversité dans divers types de contenu, notamment les émissions, les médias sociaux, les longs métrages et la radio¹⁸. Elle extrait et analyse les données en temps réel, et fournit des informations sur les paramètres de diversité tels que le genre, l'appartenance ethnique et l'âge.

Les capacités avancées de l'IA de DiversityCatch lui permettent de traiter et d'analyser rapidement de grands volumes de contenu, en surpassant les méthodes traditionnelles de collecte de données humaines. Les producteurs peuvent ainsi élaborer des stratégies pour la création de contenus plus inclusifs. Le logiciel est actuellement utilisé par des acteurs majeurs de l'industrie, notamment Netflix, des radiodiffuseurs danois et l'Union européenne de radio-télévision.

Compte tenu de la demande croissante de contenus diversifiés et des lacunes existantes en matière de données, DiversityCatch offre une solution précieuse pour promouvoir l'inclusivité dans le paysage médiatique.

1.3.3. Étude de cas 3 : Midjourney et DALL.E, outils d'IA pour la création d'images et de vidéos

Si certains outils d'IA peuvent aider à créer des images à des fins de marketing, d'autres peuvent aussi générer des vidéos comportant une intrigue. Des outils d'IA comme Midjourney¹⁹ et DALL.E²⁰ peuvent aider à concevoir des affiches de film ou à transformer des scènes de film existantes en animations. Midjourney a réalisé le premier court-métrage généré par l'IA, « *In search of time*²¹ ».

Pendant, la production d'images de haute qualité nécessite la maîtrise de techniques précises de *prompts*.

En outre, la gestion des droits liés à l'exploitation des images générées par IA suscite des inquiétudes, car leur cadre juridique reste incertain à cette heure. Les

¹⁸ <https://mediacatch.io/solution/diversitycatch>.

¹⁹ <https://www.midjourney.com/showcase>.

²⁰ <https://openai.com/index/dall-e-3/>.

²¹ <https://tribecafilm.com/films/in-search-of-time-2023>.



producteurs qui utilisent ces images pourraient s'exposer à des procédures d'infraction en raison de l'ambiguïté juridique qui entoure le contenu généré par IA.

1.4. Défis posés à l'industrie audiovisuelle par l'IA

L'essor de l'IA générative a mis en lumière les défis auxquels est confrontée l'industrie audiovisuelle (sous-section 1). Bien qu'une série de textes législatifs, qui encadrent déjà l'utilisation de l'IA en dehors et au sein de l'industrie audiovisuelle, puissent aider à résoudre certaines questions, le paysage réglementaire semble fragmenté (sous-section 2).

1.4.1. Quels sont les défis à venir pour le secteur audiovisuel ?

L'intégration de l'IA dans l'industrie audiovisuelle s'accompagne d'une myriade de défis qu'il convient d'examiner attentivement²². Certaines associations représentant l'industrie audiovisuelle ont fait part de leurs préoccupations concernant les développements de l'IA²³, mais quels sont les principaux défis que l'IA pose au secteur ? Il s'agit par exemple de :



L'impact négatif sur l'emploi : l'IA menace de bouleverser les métiers traditionnels de l'industrie audiovisuelle, ce qui pourrait entraîner des pertes d'emplois pour des professionnels tels que les acteurs vocaux et les équipes de production. Ce phénomène entraîne des répercussions sur les revenus, mais il soulève également des inquiétudes quant à la perte de créativité et de diversité au sein de la profession.



La préservation de la dimension humaine de la créativité : bien que l'IA puisse améliorer l'efficacité des processus de production et d'édition, il est nécessaire de

²² Pour en savoir plus sur les différents défis, voir: i) les résultats d'une enquête menée auprès des membres du Comité consultatif de l'OEA en mars 2024 ; ii) [l'enquête du DACS sur les artistes et l'IA, « AI and artists' work »](#), DACS, 18 janvier 2024 ; iii) [« AI is transforming the entertainment business »](#), *The Economist*, 4 janvier 2024 ; iv) Déclaration de la Société des auteurs audiovisuels, [« L'IA doit être au service de la société et renforcer la créativité humaine »](#) 4 octobre 2023 ; v) [« The impact of AI technologies on the writing profession »](#), *The Authors Guild* ; et vi) [« The AI data scraping challenge: how can we proceed responsibly ? »](#), OECD.AI, Lee Tiedrich, 5 mars 2024.

²³ Pour en savoir plus sur les différentes préoccupations soulevées par les associations, voir aussi i) SAA, [« Règlement de l'UE sur l'IA : déclaration commune des créateurs et détenteurs de droits européens »](#), document d'orientation publié le 13 mars 2024, ii) ACT, [« Réponse à l'appel à contribution de la Commission européenne sur la concurrence dans les mondes virtuels et l'IA générative »](#), document d'orientation publié le 15 mars 2024, iii) FERA, [« Déclaration commune des organisations d'auteurs, d'interprètes et autres professionnels de la création sur l'IA générative et le règlement de l'UE sur l'IA »](#), document d'orientation publié le 25 avril 2024, et iv) UER, [« L'UER salue la publication de l'AI Act par le Parlement européen »](#), document d'orientation publié le 13 mars 2024.



préserver la dimension et la créativité humaines qui font partie intégrante des processus artistiques. Des questions se posent quant à l'équilibre entre assistance de l'IA et créativité humaine, en particulier dans le contexte du financement et du soutien des institutions publiques.



Les problèmes de concurrence : la plupart des outils d'IA sur le marché sont développés et basés aux États-Unis. Leur conception n'est pas du ressort de l'UE, et l'industrie audiovisuelle européenne risque de manquer des moyens d'action nécessaires pour faire valoir ses droits de l'autre côté de l'Atlantique.



Les données utilisées et les droits d'auteur : l'utilisation de données protégées par des droits d'auteur pour entraîner des modèles d'IA sans le consentement explicite des détenteurs de droits pose des problèmes juridiques et éthiques. En outre, l'extraction de données à partir de d'internet pour la création de contenu ne va pas sans poser de questions en matière de protection des données et de respect de la vie privée.



Les droits de la personnalité : l'extraction de données soulève des questions relatives aux droits de la personnalité, car des photos, des voix ou des vidéos pourraient être utilisées pour créer du contenu généré par l'IA.



L'impact sur les salles de rédaction : l'utilisation d'outils d'IA générative dans les salles de rédaction soulève des questions sur l'intégrité journalistique et la collaboration des salles de rédaction avec les entreprises d'IA. La question du maintien d'un journalisme centré sur l'humain et du pluralisme des médias nécessite une réflexion approfondie.



La désinformation : la prolifération des contenus générés par IA suscite des inquiétudes relatives à la diffusion de désinformation et de fausses informations, et remet en cause la crédibilité des sources médiatiques et la confiance du public.



Le coût environnemental : la dépendance croissante à l'égard des technologies d'IA a des implications environnementales, notamment en termes de consommation d'énergie et de production de déchets électroniques, qui doivent être prises en compte dans le cadre du développement durable.



Les dilemmes éthiques : tous les éléments ci-dessus posent des problèmes d'ordre éthique. Il s'agit par exemple des implications du recours aux acteurs générés par l'IA pour l'industrie, y compris les questions relatives à leurs droits, à la perception du public et à l'avenir du cinéma. On peut également s'interroger sur les implications culturelles du contenu généré par l'IA au regard de la créativité humaine, son impact sur le concept de diversité culturelle et son incidence sur la démocratie dans le secteur audiovisuel. Le rôle de l'IA dans le journalisme et son impact potentiel sur les médias d'informations, notamment en ce qui concerne l'équilibre entre automatisation et dimension humaine dans les reportages, est un autre thème méritant d'être exploré.

En vue de déterminer si la législation présentée dans la section suivante 1.4.2. répondra à ces questions, les chapitres suivants (de 2 à 10) se pencheront sur ces défis et chercheront à déterminer si les cadres réglementaires sont à la hauteur de la capacité de l'IA à se développer, et ainsi à même de répondre à l'évolution des paysages technologiques au sein de l'industrie audiovisuelle.



1.4.2. Le cadre législatif de l'IA : un puzzle complexe

La législation européenne relative à l'IA forme un cadre complexe et interconnecté, où chaque élément influence et complète les autres. Elle reflète la nature multidimensionnelle des impacts et des défis que pose l'IA.

La directive sur la responsabilité du fait des produits défectueux, initialement adoptée en 1985²⁴, est en cours de révision pour tenir compte des progrès de l'intelligence artificielle. La proposition de la Commission européenne, dévoilée le 28 septembre 2022, souligne la nécessité de ces mises à jour²⁵. Parallèlement à cette révision, la Directive sur la responsabilité en matière d'IA a été proposée pour traiter spécifiquement des questions de responsabilité propres aux systèmes d'IA²⁶. Malgré ses enjeux cruciaux, le processus législatif relatif à la Directive sur la responsabilité en matière d'IA a été lent et n'a guère progressé depuis la désignation, en octobre 2022, de la commission JURI du Parlement européen.

Le Règlement sur l'IA, officiellement approuvé par le Conseil de l'UE le 21 mai 2024, impose des obligations de transparence aux fournisseurs de modèles d'IA à usage général²⁷. En outre, ils doivent veiller au respect des lois de l'Union sur le droit d'auteur, comme le prévoit l'article 53, paragraphe 1, du Règlement sur l'IA. Le texte fait référence à la directive sur le droit d'auteur dans le marché unique numérique et exige que les fournisseurs d'IA respectent les droits des créateurs de contenu, en particulier dans les scénarios impliquant la fouille de textes et de données^{28 29}.

La fouille de données est essentielle au développement de l'IA, mais elle doit se conformer à plusieurs réglementations en matière de protection des données. Le Règlement général sur la protection des données (RGPD), promulgué en 2016, fixe les

²⁴ [Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux.](#)

²⁵ Proposition de la Commission européenne de [directive en matière de responsabilité du fait des produits défectueux](#) (28 septembre 2022).

²⁶ [Proposition de directive relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle \(Directive sur la responsabilité en matière d'IA\)](#), 28 septembre 2022.

²⁷ Un modèle d'IA à usage général (*general-purpose AI*, GPAI) est un modèle d'IA qui, y compris lorsque ce modèle est entraîné à l'aide d'un grand nombre de données utilisant l'auto-supervision à grande échelle, présente une généralité significative et est capable d'exécuter de manière compétente un large éventail de tâches distinctes, indépendamment de la manière dont le modèle est mis sur le marché, et qui peut être intégré dans une variété de systèmes ou d'applications en aval, à l'exception des modèles d'IA utilisés pour des activités de recherche, de développement ou de prototypage avant leur mise sur le marché (article 3, paragraphe 63, du Règlement sur l'IA).

²⁸ [Directive \(UE\) 2019/790 sur le droit d'auteur et les droits voisins dans le marché unique numérique](#), 17 avril 2019. Selon son article 2, paragraphe 2, la fouille de textes et de données inclut toute technique d'analyse automatisée visant à analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations.

²⁹ Voir les chapitres 3 et 4 de cette publication.



bases de la protection des données dans l'ensemble de l'UE³⁰. Cette loi a été suivie par le Règlement sur la gouvernance des données de 2022, qui souligne le rôle déterminant des données dans le développement rapide des technologies de l'intelligence artificielle (considérant 2)³¹. Plus récemment, le Règlement sur les données de 2023, bien qu'il ne soit pas exclusivement lié à l'IA, a une incidence sur l'utilisation des données par les systèmes d'IA (par exemple, ceux qui impliquent des dispositifs d'internet des objets basés sur l'IA³²). Ces réglementations garantissent collectivement que le traitement et l'utilisation des données par les applications d'IA respectent les normes de protection de la vie privée et des données³³.

Lorsque le traitement des données devient une infrastructure essentielle, le droit de la concurrence (par exemple l'article 102 du TFUE³⁴) peut empêcher les entreprises dominantes d'abuser de leur pouvoir en retenant le contrôle de cette infrastructure cruciale au sein du marché intérieur de l'UE. Le droit de la concurrence comprend désormais le Règlement sur les marchés numériques (*Digital Markets Act, DMA*)³⁵, qui fait partie du paquet « services numériques » au même titre que le règlement sur les services numériques (*Digital Services Act, DSA*)³⁶. Le Règlement sur les marchés numériques régit spécifiquement la manière dont les « contrôleurs d'accès » désignés gèrent les données, une ressource vitale pour les systèmes d'IA (article 5). En revanche, le Règlement sur les services numériques exige des fournisseurs de très grandes plateformes en ligne (VLOP) qu'ils fassent preuve de transparence en matière d'algorithmes et qu'ils rendent des comptes (voir, par exemple, l'article 33).

Des réglementations spécifiques sont en place pour le secteur audiovisuel en tant que tel. La directive sur les services de médias audiovisuels fournit un cadre réglementaire pour le contenu audiovisuel, garantissant la diversité et l'équité³⁷. Le Règlement européen sur la liberté des médias (EMFA) récemment adopté comprend des dispositions relatives aux VLOP, qui les obligent à déclarer que le contenu généré par l'IA a fait l'objet d'un examen par un être humain ou d'un contrôle éditorial (article 18, paragraphe 1, point e)³⁸. Ces mesures visent à préserver l'intégrité et la qualité des contenus audiovisuels à l'ère de l'IA.

³⁰ [Règlement \(UE\) 2016/679 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 \(Règlement général sur la protection des données\)](#), 27 avril 2016.

³¹ [Règlement \(UE\) 2022/868 portant sur la gouvernance européenne des données](#), 30 mai 2022.

³² [Règlement \(UE\) 2023/2854 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données](#), 13 décembre 2023.

³³ Voir les chapitres 2 et 5 de cette publication.

³⁴ [Traité sur le fonctionnement de l'Union européenne \(TFUE, Article 102\)](#).

³⁵ [Règlement \(UE\) 2022/1925 du Parlement européen et du Conseil relatif aux marchés contestables et équitables dans le secteur numérique](#), 14 septembre 2022.

³⁶ [Règlement \(UE\) 2022/2065 relatif à un marché unique des services numériques](#), 19 octobre 2022.

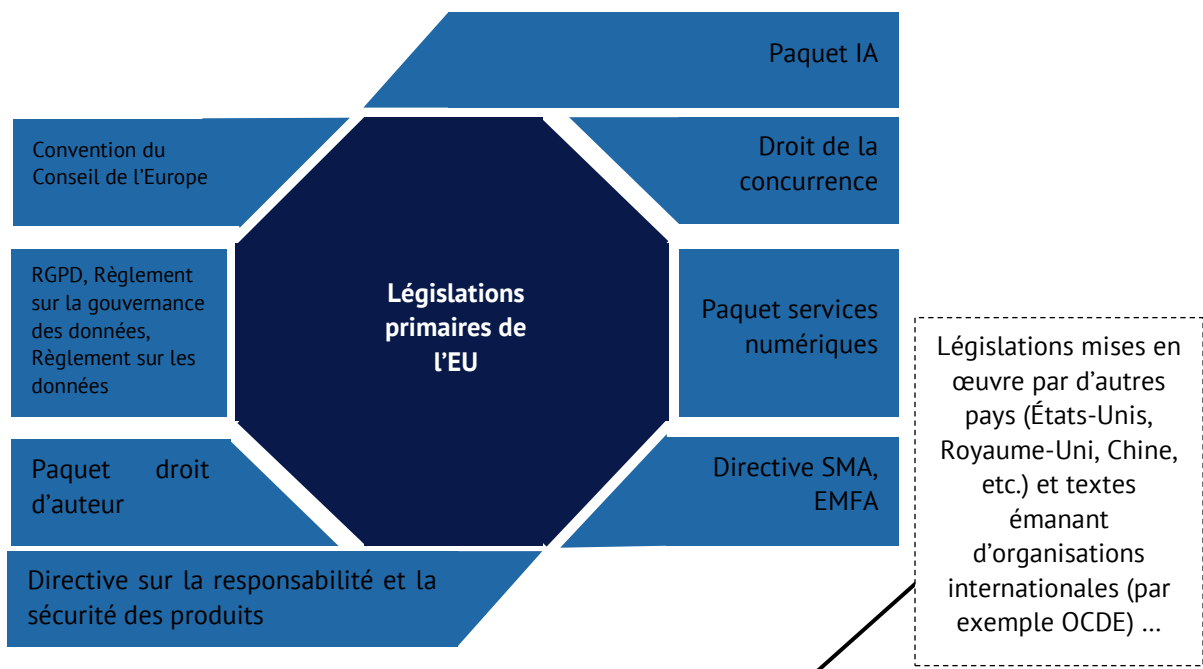
³⁷ [Directive 2010/13/UE visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à la fourniture de services de médias audiovisuels \(Directive Services de médias audiovisuels\)](#), 10 mars 2010, amendée en 2018.

³⁸ [Règlement \(UE\) 2024/1083 établissant un cadre commun pour les services de médias dans le marché intérieur](#), 11 avril 2024.



Au-delà de la réglementation européenne, certains instruments internationaux jouent un rôle crucial. La Convention du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit veille à ce que le développement et le déploiement de l'IA respectent les droits fondamentaux de l'homme et les valeurs démocratiques³⁹. Cette convention, dont l'ouverture à la signature est prévue le 5 septembre 2024, souligne la dimension mondiale de la gouvernance de l'IA et la nécessité d'une coopération internationale⁴⁰.

Schéma 3 L'IA : exemple de législations variées



Source : Observatoire européen de l'audiovisuel

L'hétérogénéité des outils législatifs, qui englobent à la fois des directives et des règlements, implique que les États membres de l'Union européenne peuvent choisir des méthodes différentes pour atteindre les objectifs des directives. Par exemple, en février 2024, la proposition de la Pologne pour la transposition de la dernière directive sur le droit d'auteur incluait une exclusion des créations de l'IA générative du champ d'application de l'exception relative à la fouille de textes et de données⁴¹.

En outre, la législation n'est véritablement mise à l'épreuve que lorsqu'elle est en application et qu'elle parvient à s'adapter aux évolutions en cours, comme l'ont montré

³⁹ [Convention-cadre sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit](#), adoptée le 17 mai 2024 par le Comité des Ministres du Conseil de l'Europe.

⁴⁰ Voir le chapitre 9 de cette publication.

⁴¹ « [TDM : Poland challenges the rule of EU copyright law](#) », *Kluwer Copyright Blog*, 20 février 2024.



les récents développements qui seront examinés plus en détail dans les prochains chapitres.

Il ne fait aucun doute que l'avenir verra se multiplier les scénarios, tant bénéfiques que problématiques pour l'industrie, qui nécessiteront la mise en place de cadres législatifs clairs dans le monde entier.

Part II – IA générative et transparence des données

« La conception de prompts est une compétence au potentiel extraordinaire », a déclaré le PDG d'OpenAI, Sam Altman, en 2023⁴².

La formulation des prompts influence directement la qualité des résultats obtenus. Alors que la conception de prompts est de plus en plus reconnue comme une nouvelle compétence de pointe, certains affirment qu'il n'en est rien, car l'IA est de plus en plus capable de comprendre le langage naturel sans avoir recours à des prompts élaborés de façon méticuleuse⁴³. Cependant, il ne faut pas oublier que les prompts activent un système entraîné à partir d'une grande quantité de données.

L'un des défis de l'IA générative ouverte est le manque de transparence de ces données. Les utilisateurs ignorent souvent les sources des données utilisées pour entraîner les machines.

Peu d'informations sont disponibles sur l'origine des données lorsqu'on utilise l'IA générative, et notamment sur la question de savoir si ces données sont protégées. Par exemple, le moissonnage de données vocales pourrait entraîner l'application de réglementations sur la protection des données telles que le RGPD.

Les préoccupations relatives au droit d'auteur sont critiques lorsqu'il s'agit de former l'IA générative à soutenir la créativité dans l'industrie audiovisuelle, qui peut être abondante en œuvres protégées par le droit d'auteur. Les données sont un nouvel or pour l'apprentissage de l'IA, mais elles pourraient aussi être une source de revenus pour les titulaires de droits. Sans transparence et divulgation suffisantes des sources de données, les détenteurs de droits pourraient ne pas être en mesure de suivre l'utilisation de leurs œuvres, de donner leur consentement ou de percevoir des droits.

⁴² <https://x.com/sama/status/1627796054040285184>

⁴³ [AI Prompt Engineering isn't the Future, Oguz A. Acar, Harvard Business Review, 6 juin 2023](#)



2. L'IA et la protection des données dans les médias audiovisuels

Philipp Hacker, LL.M. (Université de Yale)

2.1. Introduction

L'IA a un impact considérable sur le secteur audiovisuel où elle modifie profondément les processus de création, de distribution et de personnalisation des contenus. L'IA générative (GenAI), en particulier, utilise des images, vidéos et documents audio existants – souvent extraits d'Internet – pour créer des contenus audiovisuels. Or, cette révolution technologique pose des problèmes majeurs en matière de protection des données, qui doivent impérativement être traités pour garantir la conformité de l'utilisation de l'IA avec les réglementations existantes et protéger la vie privée des individus.

Il va sans dire que les données constituent la pierre angulaire du développement de l'IA, en particulier dans le secteur audiovisuel. Les technologies d'IA s'appuient principalement sur de grands ensembles de données pour entraîner des modèles qui alimentent les systèmes de recommandation, automatisent la modération des contenus et analysent les comportements du public. Dans le secteur audiovisuel, l'entraînement de l'IA est alimenté par divers types de données, dépassant les contenus protégés par le droit d'auteur pour inclure des données brutes ou traitées, des métadonnées, des contenus générés par les utilisateurs et utilisatrices, et des contenus du domaine public. Cette diversité de données permet aux systèmes d'IA d'apprendre et de s'adapter à différents contextes. Mais elle contribue également à la prolifération de discours mensongers, de préjugés et d'informations couvertes par les dispositions légales de protection des données.

En réponse à ce phénomène, un vaste dispositif réglementaire s'est développé dans le secteur audiovisuel pour relever les défis concernant la protection des données. Les réglementations pertinentes incluent le Règlement général sur la protection des données (RGPD) de l'UE, le règlement de l'UE sur l'IA récemment adopté, et d'autres dispositifs applicables hors de l'UE, par exemple aux États-Unis, au Royaume-Uni et au niveau international.

Le processus d'apprentissage automatique dans le secteur audiovisuel comprend plusieurs étapes, chacune présentant des défis distincts en matière de protection des données :



- Ensembles de données : d'importants ensembles de données sont indispensables pour l'entraînement des modèles d'IA, mais ont de lourdes implications sur la confidentialité des données. La collecte, le stockage et l'utilisation à grande échelle de données à caractère personnel doivent être gérés avec diligence pour éviter toute violation de la protection des données – ce qui peut néanmoins s'avérer impossible dans certains cas.
- Entraînement : la base juridique de l'entraînement de l'IA doit être clairement définie et les dispositions protégeant spécifiquement les données sensibles doivent être respectées.
- Modèle : une fois entraînés, les modèles d'IA doivent faire face à certains problèmes tels que l'inversion de modèle et la fuite de données, qui peuvent compromettre la confidentialité des données à caractère personnel. Le droit à l'effacement en vertu du RGPD est également crucial, car il permet aux citoyens et citoyennes de demander la suppression de leurs données des systèmes d'IA – voire même, dans certains cas, la suppression de l'intégralité du modèle.
- Déploiement : pendant le déploiement, les systèmes d'IA doivent respecter les exigences légales en matière de traitement des données, de garantie de l'exactitude des résultats, et de prévention de la diffusion de désinformation ou d'« hallucinations »⁴⁴. De plus, l'utilisation de l'IA pour la prise de décision automatisée doit tenir compte des dispositions en matière de transparence et des interdictions spécifiquement prévues. Par ailleurs, la protection des mineurs et d'autres groupes vulnérables de la population demeure une préoccupation majeure.

Tous ces éléments soulignent la relation complexe entre le développement de l'IA et la protection des données dans le secteur audiovisuel et mettent en évidence les tensions existantes entre, d'une part, un environnement technologique en pleine mutation, en particulier depuis l'avènement de GenAI, et d'autre part, les obligations légales axées sur la limitation des finalités, la minimisation des données et la limitation des durées de conservation.

2.2. Les contenus audiovisuels en tant que données à caractère personnel

Les données audiovisuelles, telles que les images, les vidéos et les enregistrements vocaux, sont considérées comme des données à caractère personnel en vertu du RGPD dès lors qu'elles se réfèrent à une personne physique identifiée ou identifiable (art. 4 du RGPD). Dans des conditions analogues, elles sont considérées comme des informations

⁴⁴ Il s'agit d'informations absurdes ou non conformes à la source fournie, voir par. 2.3. ci-après.



personnellement identifiables par d'autres dispositifs de protection des données, tels que la réglementation applicable aux États-Unis⁴⁵.

Par conséquent, les photographies et les séquences vidéo entrent dans la catégorie des données à caractère personnel dès lors qu'elles permettent d'identifier un individu. L'autorité italienne de protection des données a statué dans le même sens sur les photographies dans l'ordonnance rendue à l'encontre de Clearview AI⁴⁶. Conformément aux déclarations de l'Information Commissioner's Office (Commissariat à l'information du Royaume-Uni), si une image ou une vidéo montre, par exemple, le visage d'une personne ou d'autres caractéristiques identifiables, elle est généralement considérée comme une donnée à caractère personnel⁴⁷. Comme le note le considérant 51, lorsque ces images ou vidéos subissent un traitement technique spécifique, comme pour la reconnaissance faciale, elles peuvent même entrer dans la catégorie des données biométriques, qui sont spécifiquement protégées par l'article 9 du RGPD. Comme l'a souligné la Data Protection Commission (commission irlandaise de protection des données), une fois que les photos sont partagées en ligne, l'exemption du champ d'application du RGPD au titre d'activité domestique, qui couvre certaines activités de traitement d'ordre privé (article 2, par. 2, point c), ne s'applique plus⁴⁸.

Les enregistrements vocaux sont généralement considérés comme des données à caractère personnel, car une personne peut être identifiée par les caractéristiques vocales qui lui sont propres⁴⁹, et peuvent même constituer des données sensibles, car elles révèlent certaines caractéristiques telles que le sexe ou l'âge⁵⁰.

Les méthodes avancées d'anonymisation vocale de l'orateur visent à effacer l'identité de l'orateur⁵¹. Le procédé initial s'appuie sur des techniques de transformation vocale qui modifient la source ou filtrent certaines caractéristiques du message vocal⁵². Des recherches récentes ont étudié l'utilisation de représentations vectorielles du locuteur pour supprimer le timbre d'un locuteur ou d'une locutrice, empêchant ainsi son

⁴⁵ Voir, par exemple, Erika McCallister, Tim Grance et Karen Scarfone, *Guide to Protecting the Privacy of Personally Identifiable Information (PII)*, *Recommendations of the National Institute of Standards and Technology*, National Institute of Standards and Technology Special Publication 800-122, 2010, par. 2-1 ; Usercentrics, *Personally Identifiable Information (PII) vs. Personal Data – What's the difference?*, Usercentrics CMP, Munich, 3 mars 2021.

⁴⁶ *Ordinanza ingiunzione nei confronti di Clearview ai*, 10 février 2022, affaire 9751362, point 3.4.

⁴⁷ Information Commissioner's Office, *Taking photographs: data protection advice for schools*, Cheshire.

⁴⁸ Irish Data Protection Commission, *What is the position regarding individuals taking photographs/videos in a public place?*, Dublin.

⁴⁹ Cf. Nora Ni Loideain et Rachel Adams, « *From Alexa to Siri and the GDPR: the gendering of virtual personal assistants and the role of data protection impact assessments* », *Computer Law & Security Review* 105366, 2020, 10.

⁵⁰ Andreas Nautsch and others, *The GDPR and Speech Data: Reflections of the Legal and Technology Communities, First Steps towards a Common Understanding*, Interspeech: Crossroads of Speech and Language, 2019, p. 3.

⁵¹ Ingo Siegert et al., « *Personal data protection and academia: GDPR issues and multi-modal data-collections "in the wild"* », *Online Journal of Applied Knowledge Management*, 2020, p. 20.

⁵² Miran Pobar et Ivo Ipšić, *Online speaker de-identification using Voice transformation*, 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, 2014, p. 1264.



identification⁵³. Toutefois, il convient de garder à l'esprit qu'il existe de nombreuses techniques de ré-identification⁵⁴ qui, à terme, pourraient même être en mesure de convertir des données à caractère non personnel en données à caractère personnel (considérant 26 du RGPD)⁵⁵. D'une façon générale, la grande majorité des contenus audiovisuels sera donc considérée comme des données à caractère personnel ou des informations permettant une identification personnelle qui relèvent du champ d'application de la législation des pays respectifs en matière de protection des données.

2.3. Les problématiques en matière de protection des données et de confidentialité

Les défis à relever sont différents en fonction de la réglementation régissant la protection des données. Néanmoins, certains problèmes concernent bon nombre de législations en matière de protection des données en vigueur dans divers pays du Conseil de l'Europe, comme le montrent les dernières publications des autorités chargées de la protection des données⁵⁶. Ces problèmes portent sur les points suivants⁵⁷ :

- - la base juridique de l'inclusion de contenus audiovisuels dans un ensemble de données d'entraînement, notamment en ce qui concerne le moissonnage (*scraping*) ;
- les hallucinations et les données à caractère personnel objectivement erronées ;
- les LLM (grands modèles de langage) ;
- le traitement des données sensibles ;
- l'information et la capacité de contrôle des utilisateurs et utilisatrices ;
- la prise de décision automatisée.

⁵³ Fuming Fang *et al.*, *Speaker Anonymization Using X-vector and Neural Waveform Models*, 10th ISCA Workshop on Speech Synthesis (SSW 10), 2019.

⁵⁴ Luc Rocher, Julien M Hendrickx et Yves-Alexandre De Montjoye, « [Estimating the success of re-identifications in incomplete datasets using generative models](#) », *Nature Communications* 10, 2019, p. 1-9 ; voir aussi Paul Ohm, « [Broken promises of privacy: Responding to the surprising failure of anonymization](#) », *UCLA Law Review*, 2009, p. 1701-1777 et Manon Oostveen, « [Identifiability and the applicability of data protection to big data](#) », *International Data Privacy Law*, 2016, p. 299-309.

⁵⁵ Michèle Finck et Frank Pallas, « [They who must not be identified—distinguishing personal from non-personal data under the GDPR](#) », *International Data Privacy Law*, 2020, p. 11-36 ; Philipp Hacker et Jürgen Neyer, « [Substantively smart cities—Participation, fundamental rights and temporality](#) », *Internet Policy Review*, 2023, p. 1-30.

⁵⁶ Voir, par exemple, les lignes directrices du Comité européen de la protection des données (EDPB), *Report of the Work Against the ChatGPT Taskforce*, 23 mai 2024 ; autorités allemandes de protection des données, « [Orientierungshilfe der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder](#) », *Künstliche Intelligenz und Datenschutz*, version 1.0, 6 mai 2024 ; Office régional bavarois de protection des données, *Checkliste mit Prüfkriterien nach DS-GVO*, 24 janvier 2024 ; autorité française chargée de la protection des données (CNIL), *Guide d'auto-évaluation pour les systèmes d'intelligence artificielle (IA)* ; Information Commissioner's Office du Royaume-Uni, *Guidance on AI and Data Protection*, 15 mars 2023 ; autorité italienne de protection des données, *Instructions against web scraping*, 20 mai 2024.

⁵⁷ Voir aussi Claudio Novelli *et al.*, *Generative ai in eu Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, arXiv preprint arXiv:240107348, 2024, p. 1-36.



2.3.1. La base juridique de l'entraînement

Pour entraîner un modèle d'IA, il est nécessaire de traiter de gros volumes de contenus audiovisuels. Dès lors que ces images, vidéos ou sons constituent des données à caractère personnel (voir ci-dessus), le droit de la protection des données entre en jeu : en vertu de l'article 6 du RGPD, toute action impliquant le traitement de données à caractère personnel, tel que le moissonnage, le stockage, le transfert ou la copie, doit avoir une base juridique. Le RGPD s'applique également aux entreprises basées hors de l'UE fournissant des prestations au sein de l'UE, ce qui est le cas de nombreuses grandes entreprises d'IA. L'utilisation de données à caractère personnel pour l'entraînement de l'IA, y compris lors de la phase d'affinage, est illégale en vertu du RGPD, sauf si elle s'appuie sur une base juridique spécifique.

En général, il est impossible d'obtenir le consentement en bonne et due forme de toutes les personnes dont les données sont intégrées à de grands ensembles de données, car cela impliquerait des coûts beaucoup trop lourds⁵⁸. Par conséquent, l'entraînement de l'IA repose souvent sur le critère de mise en balance visé à l'article 6, paragraphe 1, point f), qui justifie le traitement des données si les intérêts légitimes du développeur prévalent sur les droits et libertés des personnes concernées⁵⁹. Le résultat de la mise en balance doit être évalué au cas par cas. Cependant, quelques indications générales peuvent être données.

Si un modèle d'IA a des applications d'utilité sociale ou si les personnes concernées pouvaient raisonnablement s'attendre à un traitement des données (considérant 47), la pondération peut être en faveur des développeurs. Toutefois, ce dernier critère est rarement rempli. De plus, des mesures améliorant la confidentialité comme la pseudonymisation, la transparence ou le cryptage peuvent également contribuer à la licéité de l'entraînement de l'IA. Par ailleurs, la nature et l'étendue du traitement, le type de données (en particulier les données sensibles) et le niveau de transparence et de contrôle offert aux personnes concernées peuvent jouer contre la licéité du processus⁶⁰.

Dans le cas de modèles d'IA strictement adaptés utilisant l'apprentissage supervisé, on peut considérer que l'entraînement de l'IA ne nuit pas de manière significative aux personnes concernées, en particulier si le modèle n'est pas diffusé à grande échelle et que le niveau de sécurité du système informatique prévient efficacement les violations de données⁶¹. Néanmoins, on peut difficilement reprendre cet argument pour la GenAI, car ces modèles sont généralement utilisés à grande échelle. De

⁵⁸ Miranda Mourby, Katharina Ó Cathaoir et Catherine Bjerre Collin, « [Transparency of machine-learning in healthcare: The GDPR & European Health law](#) », *Computer Law & Security Review*, 2021, 105611.

⁵⁹ Frederik J Zuiderveen Borgesius *et al*, « [Tracking walls, take-it-or-leave-it choices, the GDPR, and the ePrivacy regulation](#) », *European Data Protection Law Review*, 2017, p. 353-368.

⁶⁰ Philipp Hacker, Andreas Engel et Marco Mauer, [Regulating ChatGPT and other large Generative ai Models](#), ACM Conference on Fairness, Accountability, and Transparency (FACCT '23) 1112, Technical Report, 2023, p. 1-22.

⁶¹ Tal Z. Zarsky, [Incompatible: The GDPR in the Age of Big Data](#), 47 *Seton Hall L Rev* 995, 2016, p. 995 à 1018 ; Philipp Hacker, « [A legal framework for ai training data – from first Principles to the Artificial Intelligence Act](#) », *13 Law, innovation and Technology*, 2021, p. 257 à 301.



plus, des études récentes montrent qu'ils sont susceptibles de révéler des données à caractère personnel en cas de fuites de données ou d'attaque par inversion de modèle (voir par. 3 ci-dessous)⁶². Ce problème est encore plus grave dans les scénarios d'affinage⁶³.

Faisant écho à ces préoccupations, une récente directive restrictive de l'autorité néerlandaise de protection des données rappelle que le moissonnage en masse de données à caractère personnel sur Internet est presque toujours illicite, sauf s'il répond à une finalité spécifique bien précise⁶⁴. De plus, l'autorité italienne de protection des données a établi que le moissonnage sur Internet par Clearview AI à des fins générales de reconnaissance faciale n'avait aucun fondement juridique et ne saurait être justifié par le critère de mise en balance des intérêts⁶⁵. En général, il est difficile, voire impossible, de concilier la collecte et le traitement à grande échelle de données à caractère personnel pour les grands modèles de langage, en particulier à partir d'Internet, avec les réglementations en matière de protection des données, telles que le RGPD, qui exigent une base juridique spécifique pour le traitement des données.

2.3.2. Les hallucinations

Au-delà de l'exigence d'une base juridique, la législation en matière de protection des données définit généralement une série de principes régissant le traitement des données à caractère personnel. Comme cela a été souligné à plusieurs reprises⁶⁶, l'analyse de données à grande échelle et l'IA ne sont pas facilement conciliables avec des principes tels que la limitation des finalités, la limitation des durées de conservation ou la minimisation des données. Par ailleurs, le principe d'exactitude des données a pris une acuité particulière avec l'avènement de la GenAI et cette exigence est inscrite dans le RGPD de l'UE, mais aussi dans le RGPD du Royaume-Uni⁶⁷. Dans le secteur de l'audiovisuel, les résumés de films générés par l'IA peuvent fournir des informations inexactes sur les interprètes et les cinéastes, des hypertrucages peuvent présenter certains actes ou propos imputables aux personnes concernées alors qu'elles n'ont jamais

⁶² Voir, par exemple, Stella Biderman *et al.*, *Emergent and Predictable Memorization in large Language models*, 36 Advances in Neural Information Processing Systems, 2024, p. 1-9 ; Nicholas Carlini *et al.*, *Quantifying Memorization Across Neural Language models*, The Eleventh International Conference on Learning Representations, 2023, p. 1-19 ; Nicholas Carlini *et al.*, *Extracting training data from large language models*, 30th USENIX Security Symposium (USENIX Security 21) 2633, 2021, p. 1-13 ; Eric Lehman *et al.*, *Does BERT pretrained on clinical notes reveal sensitive data?*, ArXiv preprint arXiv:210407762, 2021, p. 1-10 ; Nicholas Carlini *et al.*, *Extract Training Data from diffusion Models* (2023) arXiv preprint arXiv:230113188, 2023, p. 1-16.

⁶³ Jaydeep Borkar, *What can we learn from data leakage and unlearning for law?*, arXiv preprint arXiv:230710476, 2023, p. 1-3.

⁶⁴ Autoriteit Persoonsgegevens, *AP: scraping bijna altijd illegaal*, 1^{er} mai 2024, p. 3-26.

⁶⁵ Garante per la protezione dei dati personali, *Ordinanza ingiunzione nei confronti di Clearview AI*, point 3,6.2, 10 février 2022, p. 1-30.

⁶⁶ Voir, par exemple, Zarsky, *Incompatible: The GDPR in the age of big data* ; Novelli *et al.*, *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, 14.

⁶⁷ ICO, *Guidance on AI and Data Protection*, 2023, p. 38.



agi ni parlé de la sorte. D'une façon générale, du fait qu'elle s'appuie sur des méthodes probabilistes, la GenAI est sujette à des hallucinations (contenu factuellement erroné, absurde ou non conforme à la source fournie⁶⁸). Même si de nouveaux outils sont à l'étude pour détecter les hallucinations⁶⁹, ils suivent eux-mêmes une approche probabiliste et il est peu probable qu'ils soient en mesure de détecter et supprimer toutes les hallucinations face à des scénarios critiques⁷⁰.

Par conséquent, bien que le principe de l'exactitude des données soit crucial, il peut être mis en balance avec d'autres droits. Dans la pratique, seules les fausses informations d'une importance majeure sont susceptibles de donner lieu à une rectification⁷¹. Or, même en limitant la prévention à cette catégorie restreinte d'hallucinations, le défi reste entier pour les développeurs et les déployeurs de LLM⁷².

2.3.3. Les LLM en tant que données à caractère personnel

La réglementation actuelle en matière de protection des données, comme le RGPD, prévoit un droit à l'effacement des données à caractère personnel, ce qui devient plus complexe et problématique avec l'IA en cas d'inversion des modèles et de fuites de données. L'inversion de modèle permet de reconstruire les données d'apprentissage, y compris les contenus audiovisuels protégés, et la mémorisation peut amener l'IA à fournir des données à caractère personnel incluses dans les données d'apprentissage, même en réponse à de simples requêtes. Cela implique que les LLM peuvent être eux-mêmes considérés comme des données à caractère personnel. En ce cas, la simple mise à jour ou le simple téléchargement de LLM requiert une base juridique, et les personnes concernées pourraient être habilitées à demander la suppression du modèle en vertu de l'article 17 du RGPD. Si les LLM sont effectivement classés comme des données à caractère personnel, cela pourrait signifier une avalanche de violations de la protection des données par les entités développant ou utilisant ces modèles.

Une récente directive de l'autorité de protection des données de Hambourg du 15 juillet 2024 vise à rassurer les utilisateurs et utilisatrices sur le fait que les LLM ne sont généralement pas considérés comme des données à caractère personnel⁷³. Toutefois, cette décision ne met pas fin au débat⁷⁴. Les LLM sont comparables à des données compressées

⁶⁸ Voir uniquement Ziwei Ji *et al.*, *Survey of hallucination in Natural language generation*, ACM Computing Surveys, 2023, p. 1-3.

⁶⁹ Sebastian Farquhar *et al.*, « *Detecting hallucinations in large language models using semantic entropy* », *Nature*, 2024, p. 625-630.

⁷⁰ Cf. *ibid.*, 629.

⁷¹ Cf. de nouveau ICO, *Guidance on AI and Data Protection*, 2023, 39.

⁷² Cf. également *EDPB Report*, par. 29 à 31.

⁷³ <https://datenschutz-hamburg.de/news/hamburger-thesen-zum-personenbezug-in-large-language-models>.

⁷⁴ En faveur de la classification des LLM comme des données à caractère personnel, voir par ex. Michael Veale, Reuben Binns et Lilian Edwards, « *Algorithms that remember: model inversion attacks and data protection law* », *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2018, 20180083 ; Paulina JO Pesch et Rainer Böhme, « *Verarbeitung personenbezogener Daten und Datenrichtigkeit bei großen Sprachmodellen* », *Multimedia und Recht*, 2023, p. 920 ; contre cette conception, voir par ex.



et cryptées ; par conséquent, ces données peuvent être à caractère personnel si certaines conditions sont remplies : cela dépend de la capacité technique à faire le lien entre le modèle et des individus spécifiques, de la probabilité que le ou la responsable du traitement ait recours à cette méthode, et du débat juridique en cours sur l'impact de la licéité de la méthode sur une telle classification⁷⁵.

2.3.4. Les données sensibles

Un autre défi à caractère urgent dans le cadre du droit de la protection des données porte sur la capacité des contenus audiovisuels à révéler des informations sensibles telles que l'âge, l'origine raciale ou ethnique, les opinions politiques, les croyances religieuses ou philosophiques et l'appartenance syndicale ; cela peut être le cas, par exemple, avec des photographies (âge, origine raciale ou ethnique, religion) ou même des enregistrements vocaux (âge)⁷⁶. Ce problème a été traité notamment dans l'affaire *Meta contre Bundeskartellamt*, dans laquelle le tribunal a statué que les données n'avaient pas besoin de se référer directement à des critères sensibles pour être protégées en vertu de l'article 9 du RGPD. Il suffit que le « traitement de données [permette] de révéler des informations relevant d'une de ces catégories »⁷⁷. Or, avec des techniques d'analyse avancées, ce sera souvent le cas. Par exemple, le moteur de recommandation de contenus du module d'IA peut, délibérément ou par inadvertance, traiter des données sensibles dans ce cadre, telles que des informations sur l'âge, l'origine ethnique, la religion ou les opinions politiques des destinataires des recommandations. De plus, les données biométriques, telles que les images ou les vidéos utilisées à des fins d'identification dans la reconnaissance faciale, relèvent également de l'article 9 du RGPD⁷⁸.

L'article 9, paragraphe 2 du RGPD énonce certaines exceptions pour le traitement des données sensibles, mais ces exceptions sont limitées. En vertu de l'article 9, paragraphe 2, point e), lorsque les données sont « manifestement rendues publiques par la personne concernée », cela constitue une exception. Toutefois, la publication volontaire par la personne concernée ne légitime pas l'utilisation des données à des fins autres que l'intention initiale de la publication⁷⁹. Dans sa décision contre Clearview AI,

Flemming Moos, « [Personenbezug von Large Language Models](#) », *Computer und Recht*, 2024, par. 27 et suiv. ; voir également [EDPB Report](#), par. 25.

⁷⁵ Voir l'affaire [Patrick Breyer](#), arrêt du 19 octobre 2016, C-582/14.

⁷⁶ Voir [Ordinanza ingiunzione nei confronti di Clearview AI](#), ordonnance du 10 février 2022, affaire 9751362, point 3.4.

⁷⁷ [Meta Platforms et al.](#), arrêt du 4 juillet 2023, C-252/21, par. 73.

⁷⁸ Considérant 51 du RGPD et [Ordinanza ingiunzione nei confronti di Clearview AI](#), ordonnance du 10 février 2022, affaire 9751362, point 3.4.

⁷⁹ Groupe de travail « Article 29 » sur la protection des données, [Avis 06/2014 sur la notion d'intérêt légitime poursuivi par le responsable du traitement des données au sens de l'article 7 de la directive 95/46/CE](#) du 9 avril 2014, 39 ; [Ordinanza ingiunzione nei confronti di Clearview AI](#), ordonnance du 10 février 2022, affaire 9751362, point 3.4, « De même, il est à noter que la publication sur Internet de données à caractère personnel par la personne à laquelle elles se réfèrent, par exemple dans le cadre d'un réseau social, ne constitue pas en soi une condition suffisante pour légitimer leur libre réutilisation par des tiers. » [traduction automatique].



l'autorité italienne de protection des données établit qu'aucune exception ne saurait s'appliquer au moissonnage systématique d'images sur Internet à des fins de reconnaissance faciale, même si ces images ont été publiées volontairement par les personnes concernées⁸⁰.

Par conséquent, hormis le consentement explicite, qui est difficile à obtenir, il n'existe pas d'exception claire pour l'utilisation de données sensibles dans des modèles génératifs généraux et de contenus audiovisuels. Dans certains contextes, tels que les situations liées à la santé, des exceptions spécifiques peuvent être inscrites dans le droit national en étant assorties de garanties appropriées. Toutefois, ces exceptions sont définies de manière stricte et ne s'appliquent pas de manière générale aux modèles d'IA générative ni au traitement de contenus audiovisuels.

2.3.5. Information et capacité de contrôle des utilisateurs et utilisatrices

Les prochains défis importants pour assurer la conformité des LLM (ou autres modèles de GenAI) avec le RGPD résident principalement dans l'obligation de fournir des notifications et des informations aux personnes concernées, notamment en vertu des articles 12 à 15 du RGPD. Ces exigences posent des problèmes inédits en raison de la multitude et de la diversité des données traitées par la GenAI⁸¹.

L'article 14 du RGPD est particulièrement pertinent au regard des données collectées sur Internet à des fins d'apprentissage. Toutefois, l'obligation d'informer toute personne dont les données sont incluses dans l'ensemble de données d'apprentissage peut s'avérer irréalisable, car cela implique des efforts considérables. C'est là que l'article 14, paragraphe 5, point b) du RGPD intervient en prévoyant des exceptions lorsque l'effort est disproportionné. Conformément au considérant 62 du RGPD, les critères à prendre en compte lors de cette évaluation portent sur le nombre de personnes concernées, l'ancienneté des données, ainsi que les garanties mises en œuvre. Le Groupe de travail « Article 29 » souligne également l'impossibilité d'informer les personnes concernées lorsque les données proviennent de plusieurs sources et que les coordonnées des personnes à contacter ne sont pas disponibles⁸².

En revanche, les données à caractère personnel soumises par les utilisateurs et utilisatrices via les interfaces de *chat* (sur demande) ne bénéficient pas de telles dérogations. L'article 13 du RGPD exige explicitement que les personnes concernées disposent de plusieurs informations clés, notamment les finalités du traitement, la base

⁸⁰ [Ordinanza ingiunzione nei confronti di Clearview AI](#), ordonnance du 10 février 2022, affaire 9751362, point 3.6.3.

⁸¹ Hacker P., Engel A. et Mauer M., [Regulating ChatGPT and Other large Generative AI Models](#), ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), 5 février 2023, 2-3.

⁸² Groupe de travail « Article 29 » sur la protection des données, [Lignes directrices sur la transparence au sens du règlement \(UE\) 2016/679](#), WP260 rév. 01, Bruxelles, 2018, par. 63.



juridique du traitement et les intérêts légitimes poursuivis par le ou la responsable du traitement. Cela vaut également pour tout contenu audiovisuel que les personnes concernées peuvent téléverser.

Il est délicat de trouver l'équilibre entre la difficulté à mettre en pratique les obligations légales et le respect des droits des personnes concernées. Bien que l'article 14, paragraphe 5 du RGPD prévoit une dérogation possible en cas d'efforts disproportionnés, cela reste controversé, en particulier lorsqu'il s'agit de moissonnage et de traitement des données à des fins commerciales. Le ou la responsable du traitement des données, conformément à l'article 4, paragraphe 7 du RGPD, doit documenter minutieusement ses considérations au regard de cette disposition afin de garantir le respect du principe de responsabilité consacré à l'article 5, paragraphe 2 du RGPD. De plus, le fait de rendre publics les documents concernant les méthodes de collecte des données d'apprentissage permet de renforcer le respect des principes de protection des données tout en améliorant la transparence.

2.3.6. La prise de décision automatisée

Concrètement, l'utilisation de modèles d'IA, tels que les LLM, peut également être classée dans des processus de décision automatisés régis par le RGPD. L'article 22 interdit d'une façon générale les décisions fondées exclusivement sur un traitement automatisé (y compris le profilage) produisant des effets juridiques ou significativement similaires sur la personne concernée, sauf dans certains cas spécifiques justifiant une exception. Cette disposition est particulièrement pertinente dans des situations telles que le recrutement de personnel ou l'octroi d'un crédit, où l'évaluation automatisée peut influencer les résultats de manière significative ; cependant, les moteurs de recommandation de contenus, les hypertrucages ou les résumés de films automatisés peuvent également produire des effets significatifs. L'arrêt de la CJUE rendu récemment dans l'affaire Schufa a revu à la baisse les critères permettant de caractériser une décision automatisée⁸³ : il suffit qu'une valeur de probabilité produite par une partie (par ex. un fournisseur d'IA) détermine de façon prépondérante le fait qu'une tierce partie (par ex. un employeur, une banque ou un magasin) décide d'établir, d'exécuter ou de mettre fin à une relation contractuelle avec la personne concernée.

Les exceptions à cette interdiction sont limitées et couvrent les cas dans lesquels le consentement explicite est obtenu, le traitement est nécessaire à la conclusion ou l'exécution d'un contrat, ou certaines dispositions juridiques spécifiques s'appliquent. Toutefois, l'obtention d'un consentement valable peut s'avérer difficile en raison du déséquilibre en matière d'autorité et de pouvoir, et il est probable que les arguments fondés uniquement sur l'efficacité ne conviennent pas (considérant 43 du RGPD). En revanche, les entreprises doivent démontrer aux personnes concernées l'existence d'avantages tangibles.

⁸³ CJUE, [SCHUFA Holding \(Scoring\)](#), arrêt du 7 décembre 2023, C-634/21, par. 73.



Les affaires susmentionnées et les dispositions réglementaires illustrent une fois de plus le besoin croissant de transparence et de conformité juridique dans l'utilisation des systèmes automatisés et de l'IA pour garantir la protection des droits individuels dans un environnement de plus en plus numérique et automatisé.

2.4. Le règlement sur l'IA

Le règlement européen sur l'IA récemment adopté⁸⁴ impose plusieurs contraintes majeures aux fournisseurs et aux déployeurs d'IA lors du traitement de contenus audiovisuels, que ce soit pour l'entraînement des systèmes ou l'inférence. Il fixe un cadre global pour la gestion des risques associés au traitement de contenus audiovisuels par les systèmes d'IA. Les fournisseurs doivent mettre en œuvre de solides mesures de gestion des risques, de gouvernance des données et de transparence, tandis que les déployeurs se voient investis de responsabilités en matière de surveillance, de documentation et d'évaluation de l'impact. La transparence est renforcée par des exigences claires en matière d'information et d'étiquetage, ce qui renforce les obligations de transparence inscrites dans le RGPD⁸⁵.

Cependant, il existe également des champs de tension entre le règlement sur la protection des données et le Règlement sur l'IA⁸⁶. Le Règlement sur l'IA introduit de nouvelles fonctions et une terminologie différente avec, notamment, les « fournisseurs » (développeurs) et les « déployeurs » (utilisateurs professionnels) de systèmes d'IA, qui ne correspondent pas exactement aux fonctions de « responsable de traitement » et de « sous-traitant » du RGPD. Du fait de ces divergences, il pourrait s'avérer complexe de déterminer à l'avenir les responsabilités en matière de conformité, en particulier lorsque la même entité peut être considérée à la fois comme un simple déployeur en vertu du Règlement sur l'IA et comme un responsable de traitement en vertu du RGPD – comme ce sera souvent le cas⁸⁷. Par ailleurs, des difficultés pratiques subsistent en matière d'application et de coopération entre les différentes autorités réglementaires. Par

⁸⁴ Voir par ex. Michael Veale et Frederik Zuiderveen Borgesius, « [Demystifying the Draft eu Artificial Intelligence Act—Analysing the good, the Bad, and the unclear elements of the Proposed Approach](#) », *Computer Law Review International*, Otto Schmidt, Cologne, 2022, p. 97 ; Martin Ebers *et al.*, [The European commission's proposal for an artificial intelligence act—a critical assessment by members of the robotics and AI law society \(RAILS\)](#), MDPI, Bâle, J 2021, p. 589.

⁸⁵ Voir par ex. Philipp Hacker et Jan-Hendrik Passoth, « [Variety of AI explanations under the Law. From the GDPR to the AIA, and Beyond](#) », *xxAI – Beyond Explainable AI*, Springer, Cham, 2022, p. 343.

⁸⁶ Voir par ex. James Clark, Muhammed Demircan & Kalyna Kettas, [Europe: The EU AI Act's relationship with data protection law: key takeaways](#), Privacy Matters, DLA Piper, 25 avril 2024 ; Sergio Barezzani, « [Artificial Intelligence Act \(AI Act\) and the GDPR](#) », *Encyclopedia of Cryptography, Security and Privacy*, Springer, Cham, 2024, p. 1-6 ; Christiane Lawson-Hetchely, [The Potential Impact of the Future AI Act on the GDPR](#), Université d'Oslo, Oslo, 2022.

⁸⁷ Sebastião Barros Vale, [GDPR and the AI Act interplay: Lessons from FPF's ADM Case-Law Report](#), Future of Privacy Forum, 3 novembre 2022.



exemple, les deux règlements ont un périmètre de surveillance différent, ce qui risque d'entraîner des empiètements ou des conflits entre les mesures réglementaires⁸⁸.

2.5. Les transferts internationaux de données

D'autres obligations en matière de protection des données entrent en jeu lorsque des données audiovisuelles ou autres sont envoyées hors de l'UE, notamment pour permettre l'analyse du traitement au niveau du *cloud*. Par exemple, les radiodiffuseurs qui utilisent l'IA pour le doublage de contenus en plusieurs langues peuvent transférer des vidéos de l'UE vers un système de *cloud* basé aux États-Unis et des appareils domestiques intelligents peuvent envoyer des enregistrements vocaux à des serveurs dans des pays tiers pour le traitement du langage naturel. Les articles 44 et suivants du RGPD énoncent les règles et les garanties requises pour les transferts internationaux de données de l'UE vers des pays tiers, tels que les États-Unis. Les articles susmentionnés visent à garantir que les données à caractère personnel transférées hors de l'UE bénéficient d'un niveau de protection substantiellement équivalent à celui qui est garanti dans l'UE. De surcroît, une décision d'adéquation de la Commission européenne permet des transferts de données vers des pays réputés garantir un niveau adéquat de protection des données, simplifiant ainsi le processus de conformité pour les entités opérant à l'échelle internationale.

Le « cadre de protection des données UE-États-Unis » constitue le fondement de la dernière décision d'adéquation visant à remplacer le « bouclier de protection des données UE-États-Unis » invalidé par l'arrêt *Schrems II*⁸⁹. Cet arrêt établit que la collecte et le traitement à grande échelle de données à caractère personnel par les autorités américaines pour des raisons de sécurité nationale ne sont pas compatibles avec le principe de proportionnalité et la garantie d'un niveau adéquat de protection de la vie privée, et que les citoyens et citoyennes de l'UE ne disposent pas d'un recours juridictionnel effectif pour contester d'éventuelles violations. Dans ce contexte, le cadre de protection des données UE-États-Unis instaure des garanties renforcées, notamment des mécanismes de surveillance et d'application plus stricts, et de nouvelles voies de recours pour les citoyens et citoyennes de l'UE⁹⁰.

Le cadre de protection des données UE-États-Unis met explicitement l'accent sur les principes de nécessité et de proportionnalité pour que l'accès aux données par les autorités américaines soit strictement limité à ce qui est nécessaire et proportionné à des fins de sécurité nationale⁹¹. De plus, ce dispositif prévoit la mise en place d'une Cour chargée du contrôle de la protection des données (*Data Protection Review Court - DPRC*),

⁸⁸ Paweł Hajduk, *AI Act and GDPR: On the Path Towards Overlap of the Enforcement Structures*, RAILS Blog, RAILS, Berlin, 1^{er} octobre 2023.

⁸⁹ *Schrems II*, arrêt du 16 juillet 2020 de la CJUE, affaire C-311.

⁹⁰ Voir par ex. David Michael Watry, *The transatlantic Data Privacy framework: Schrems II, GDPR and American national security*, Université de Malte, 2023 ; Linda Kidwell, *GDPR Compliance in EU-US Data Transfers*, Université de Lund 2023.

⁹¹ Alex Wodi, *The EU General Data Protection Regulation (GDPR): Five Years After and the Future of Data Privacy Protection in Review*, document de travail, 2023, 9.



un organe indépendant et impartial qui offre aux citoyens et citoyennes de l'UE un mécanisme de recours pour traiter les réclamations concernant la collecte et l'utilisation de leurs données par les agences de renseignement américaines⁹². Cette possibilité de recours est importante, par exemple, si des journalistes d'investigation de l'UE utilisent une société d'IA basée aux États-Unis pour vérifier l'authenticité d'une vidéo montrant un événement important : ces journalistes peuvent désormais contester l'accès des services de renseignement américains à la vidéo. La DPRC est également habilitée à ordonner la suppression de données si elle constate que les données ont été collectées en violation des garanties convenues⁹³. À titre d'exemple, si un prestataire de services de post-production basé aux États-Unis (sous-titrage ou synchronisation de films à l'aide de l'IA) ne respecte pas les principes du cadre de protection des données, l'entreprise de l'UE concernée peut saisir la DPRC pour faire appliquer la réglementation.

L'impact du « cadre de protection des données UE-États-Unis » sur la conformité avec le RGPD est significatif, car il a pour vocation de répondre aux questions soulevées par la CJUE dans l'arrêt Schrems II. Néanmoins, il pourrait aussi être invalidé à terme, car le champ des possibilités de procéder au traitement de données en masse est vaste : il peut être autorisé s'il « est établi qu'il est nécessaire de procéder à la collecte en masse pour faire avancer une priorité officiellement reconnue en matière de renseignement⁹⁴. » À cet égard, la conception américaine de la notion de nécessité peut aller au-delà des exigences de stricte nécessité et de proportionnalité inscrites dans la doctrine de la CJUE⁹⁵, ce qui permet de craindre un éventuel arrêt Schrems III et de nouvelles incertitudes concernant les transferts internationaux de données entre l'UE et les États-Unis.

2.6. Comparaison avec le droit américain et le droit international

Le RGPD, la loi HIPAA et diverses lois des États américains visent tous à protéger les données à caractère personnel, mais se caractérisent par des cadres et des champs d'application différents. Le RGPD fournit une protection complète des données dans toute l'UE, établissant des garanties solides pour toutes les données à caractère personnel, avec un régime particulièrement strict pour les données sensibles, dont, notamment, les informations médicales. À titre d'exemple, toute utilisation d'éléments audiovisuels pour l'entraînement d'une IA médicale dans le cadre du RGPD doit respecter des exigences strictes en matière de transparence et, en général, de consentement. Cette disposition est similaire à la loi HIPAA aux États-Unis, qui impose la protection des données médicales.

⁹² *Ibid.*

⁹³ Commission européenne, *Questions et réponses: cadre de protection des données UE - États-Unis*, 10 juillet 2023.

⁹⁴ 50 U.S.C. § 3001, ex. Ord. No 14086, 7 octobre 2022, 87 F.R. 62283, par. 2 c) ii) A).

⁹⁵ Bjørn Aslak Juliussen *et al.*, « *The third country problem under the GDPR: enhancing protection of data transfers with technology* », *International Data Privacy Law* 2023, p. 225, 229.



Lors de l'utilisation d'images médicales pour l'entraînement d'une IA, la loi HIPAA exige l'anonymisation des données ou l'obtention du consentement explicite du patient ou de la patiente afin de garantir la confidentialité et la sécurité⁹⁶.

La législation de certains États américains, telles que la *California Consumer Privacy Act* (CCPA) ou la *Colorado Privacy Act* (CPA), reprennent de nombreux principes du RGPD en appliquant des mesures strictes de protection des données et en instaurant des droits d'accès, de suppression et de retrait du consentement au traitement des données⁹⁷. Ces réglementations prévoient des niveaux de protection supplémentaires qui s'apparentent à l'approche globale du RGPD. Par conséquent, les entreprises qui traitent des données audiovisuelles pour l'entraînement de l'IA ou à d'autres fins doivent mettre en œuvre des mesures strictes de confidentialité et, dans la mesure du possible, obtenir le consentement explicite des individus.

Toutefois, les initiatives en place au niveau des États, couplées à des mesures sectorielles au niveau fédéral (par ex. dans le cadre du *Biden Executive Order on AI* [décret-loi de Biden sur l'IA]), compliquent de plus en plus les règles de transfert de données, et une réglementation globale dans l'UE et en Chine risque de créer une juxtaposition de règles en matière de vie privée, de protection des données et d'IA applicables à l'entraînement et au déploiement de l'IA, notamment - mais pas exclusivement - dans le secteur audiovisuel.

Par conséquent, il est crucial de mener des initiatives à l'échelle internationale pour pouvoir tracer la voie à suivre dans ce labyrinthe qui ne cesse de croître. Des initiatives comme le Pacte numérique mondial des Nations Unies et le Processus du G7, dit d'Hiroshima, sur l'intelligence artificielle générative reflètent un consensus croissant sur la nécessité de garantir une IA responsable et des normes de protection des données au niveau mondial. Ces dispositifs visent à harmoniser les réglementations en matière d'IA au-delà des frontières en promouvant des principes fondamentaux, tels que la transparence, la responsabilité et la protection des droits humains, à l'instar du RGPD. Ces initiatives d'envergure mondiale sont décisives pour créer une approche cohérente de la gouvernance de l'IA, en garantissant la protection des données audiovisuelles et autres informations personnelles indépendamment de l'endroit où elles sont traitées – tout en veillant également à ce que la conformité effective reste possible pour les entreprises qui utilisent des données audiovisuelles ou autres pour le bénéfice de la société.

À terme, ces initiatives internationales devront s'articuler avec les normes internationales émergentes élaborées par des organismes normatifs tels que l'ISO ou le CEN/CENELEC, afin de concrétiser sur le terrain et dans les systèmes d'apprentissage automatique des principes aux contours encore flous. Parallèlement, cela souligne la nécessité d'associer au plus vite aux mesures de normalisation un vaste panel de parties prenantes, plus large que l'industrie, et de créer des outils efficaces, notamment avec des

⁹⁶ Steve Alder, « *Editorial: HIPAA, Healthcare Data, and Artificial Intelligence* », *The HIPAA Journal*, 16 décembre 2022 ; Becky Whittaker, « *Healthcare ai and HIPAA Privacy Concerns: Everything you need to know* », *The Intake*, 15 décembre 2022.

⁹⁷ Bloomberg Law, *Which States Have Consumer Data Privacy Laws?*, 18 mars 2024.



bourses ou divers autres dispositifs, pour impliquer la société civile et les universitaires dans la mise en place de ces mesures.



3. Entraînement de l'IA et préservation des droits d'auteur

Gianluca Campus⁹⁸, PwC Digital Innovation

3.1. Introduction

3.1.1. Aperçu des systèmes d'IA et de leur traitement des données protégées par le droit d'auteur

La nature disruptive des technologies de l'IA est évidente pour tous depuis qu'elle est capable de reproduire (et même de surpasser) les capacités humaines. Cependant, avec l'introduction de l'IA générative, de nouveaux défis juridiques cruciaux se posent du point de vue de la propriété intellectuelle.

Cette section du rapport se concentre sur le risque de violation du droit d'auteur découlant de l'utilisation d'œuvres comme données d'apprentissage pour les systèmes d'IA générative et analyse la manière dont les législateurs et les tribunaux abordent ces défis juridiques.

Tout d'abord, il est utile de comprendre comment les données d'apprentissage sont traitées dans un système d'IA générative. Pour comprendre plus en détail comment les résultats générés par l'IA dérivent des œuvres incluses dans les données d'apprentissage, il a été suggéré de prendre en compte une sorte de « chaîne d'approvisionnement de l'IA générative⁹⁹ », un ensemble interconnecté d'étapes qui

⁹⁸ Directeur des opérations juridiques chez PwC Digital Innovation Italie, docteur en droit, membre de l'université de Milan.

⁹⁹ Voir Katherine Lee, A. Feder Cooper et James Grimmelmann, « [Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain](#) », 27 juillet 2023, à paraître dans *Journal of the Copyright Society* 2024. Sur les différences majeures entre le processus créatif des systèmes d'IA et la créativité humaine et sur l'impact de ces différences sur la reconstruction des aspects du droit d'auteur, voir également Giancarlo Frosio, « Should we ban Generative AI, incentivise it or make it a medium for inclusive creativity ? », 31 juillet 2023, dans Enrico Bonadio et Caterina Sganga (éd.), [A Research Agenda for EU Copyright Law](#) (Edward Elgar, à paraître),



transforment les données d'apprentissage en créations (par exemple, une nouvelle image, que l'on espère inédite, d'un objet qui peut ou non avoir existé).

Selon la reconstruction des auteurs, la chaîne d'approvisionnement commence par les œuvres créatives : tous les livres, œuvres d'art, logiciels et autres produits de la créativité humaine dont l'IA générative s'inspire et qu'elle cherche à imiter. Ensuite, les œuvres et autres informations doivent être converties en données : des fichiers codés numériquement dans des formats standards et connus. Les données individuelles sont inutiles en elles-mêmes pour l'apprentissage de l'IA. Elles doivent être compilées dans des ensembles de données d'apprentissage : de vastes collections soigneusement structurées de données connexes. Ce processus nécessite à la fois une automatisation poussée et une prise de décision humaine réfléchie.

Pour mettre au point un modèle d'IA générative, son créateur choisit une architecture technique, regroupe des ensembles de données d'apprentissage, puis exécute un algorithme d'entraînement pour coder les caractéristiques des données d'apprentissage dans le modèle. L'entraînement d'un modèle est à la fois une science et un art, et il implique des investissements massifs en temps, en argent et en ressources informatiques. Le modèle qui résulte de ce processus d'apprentissage initial est appelé « base » ou « modèle pré-entraîné », car il s'agit souvent d'un simple point de départ. Un modèle peut également être affiné afin d'améliorer ses performances ou de l'adapter à un domaine spécifique. Ce processus implique également des choix importants, et il devrait être mené par une autre entité que celle qui a effectué la formation initiale.

Un système déployé peut être utilisé pour générer des produits : de nouvelles œuvres créatives basées sur des tendances statistiques présentes dans les données d'apprentissage, mais qui les combinent de manière inédite. Un produit – ou une « génération » - est basé sur un *prompt* donné par l'utilisateur : une entrée qui décrit les caractéristiques particulières qu'il souhaite attribuer au produit. C'est généralement la seule partie de la chaîne d'approvisionnement que les utilisateurs voient.

Dans une telle reconstruction, le modèle est simplement un arrangement différent et compliqué des exemples d'apprentissage. Mais le modèle peut également être considéré comme une œuvre dérivée de ses données d'apprentissage, une œuvre basée sur une ou plusieurs œuvres préexistantes qui combine la paternité d'une œuvre existante avec une nouvelle paternité. Les ensembles de données d'apprentissage contiennent des copies littérales complètes de millions d'œuvres numérisées protégées par le droit d'auteur. Un modèle, en tant qu'ensemble de paramètres, est différent par nature des œuvres protégées par le droit d'auteur sur lesquelles il a été formé.

selon lequel « *Un facteur qui exige un examen attentif lorsqu'on envisage des incitations juridiques pour la créativité générée par l'IA est la nature unique de la créativité générée par les machines, qui diffère considérablement des processus créatifs humains. Dans ce contexte, il est crucial de réfléchir aux caractéristiques distinctives de la créativité générée par les machines, qui excellent dans les processus cumulatifs et combinatoires [...]. Contrairement aux machines, les humains ne se souviennent pas des objets réels eux-mêmes, mais plutôt des idées conceptuelles relatives à ces objets* ».



3.1.2. Considérations sur les œuvres dérivées

Il n'est ni simple ni évident de comprendre s'il existe une relation « dérivée » entre l'ensemble de données d'apprentissage et le produit généré par l'IA. Il est essentiel de déterminer si le rendu généré par les systèmes d'IA après traitement des données peut être considéré comme une œuvre dérivée et, par conséquent, si les titulaires de droits sur les données utilisées pour l'entraînement doivent autoriser l'œuvre dérivée générée par l'IA. En ce qui concerne le système juridique américain, le professeur Daniel Gervais¹⁰⁰ souligne que la loi sur le droit d'auteur prévoit un droit exclusif « d'élaborer des œuvres dérivées basées sur l'œuvre protégée par le droit d'auteur » et définit « l'œuvre dérivée » en partie comme toute œuvre « basée sur une ou plusieurs œuvres préexistantes ». Transposé dans l'environnement de l'IA, il faut tenir compte du fait que les systèmes d'IA peuvent produire un contenu littéraire et artistique (« *output* ») qui est presque forcément « basé » sur un ensemble de données constitué d'œuvres préexistantes.

En outre, les œuvres dérivées doivent satisfaire à l'exigence d'originalité pour pouvoir bénéficier de la protection du droit d'auteur. L'« originalité » n'est pas définie par la loi, mais elle a été précisée par la Cour suprême des États-Unis comme signifiant que l'œuvre dérivée doit être créée de manière indépendante par son auteur et doit incarner une expression qui est au moins un minimum créative (c'est-à-dire que l'œuvre est le résultat de choix créatifs faits par l'auteur¹⁰¹).

En outre, la notion d'originalité appliquée à la protection des œuvres dérivées exige que la personne se présentant comme l'auteur d'une œuvre dérivée ait ajouté à, ou transformé d'une manière ou d'une autre, une ou plusieurs œuvres préexistantes. La nature juridique de l'œuvre dérivée peut découler d'une autorisation du titulaire du droit d'auteur, d'une exception telle que l'usage loyal, ou du fait que les œuvres sous-jacentes ne sont plus protégées.

Il convient donc de vérifier si les choix créatifs effectués par l'auteur du programme (ou, le cas échéant, par l'utilisateur) sont présents dans le produit du système d'IA. Si ce n'est pas le cas, la protection de ce produit en tant qu'œuvre du programmeur (ou de l'utilisateur) est incompatible avec la doctrine fondamentale du droit d'auteur et ses objectifs, et elle conférerait un avantage excessif au programmeur (ou à l'utilisateur).

En ce qui concerne le système juridique de l'UE, les principes de la Convention de Berne sont applicables, selon lesquels « sont protégés comme des œuvres originales, sans préjudice des droits de l'auteur de l'œuvre originale, les traductions, adaptations, arrangements de musique et autres transformations d'une œuvre littéraire ou artistique¹⁰² ». En outre, la CJUE a également posé une condition d'« originalité » pour les œuvres dérivées et a précisé que le critère d'originalité de l'UE exigeait plus que des compétences, du travail ou des efforts et a conclu, plus précisément, que les

¹⁰⁰ Voir Daniel J. Gervais, « [AI derivatives : the application to the derivative work right to literary and artistic productions of AI machines](#) », *Seton Hall Law Review*, volume 53, 2022 et *Vanderbilt Law*, article n°22-12.

¹⁰¹ Voir *Feist Publ'ns, Inc. c. Rural Tel. Serv. Co.*, 499 U.S. 340, 346 (1991) ; *Burrow-Giles Lithographic Co. c. Sarony*, 111 U.S. 53 (1884).

¹⁰² Voir [l'article 2\(3\) de la Convention de Berne](#).



considérations, règles et contraintes techniques ne conféraient pas un caractère original¹⁰³.

En l'absence d'indications claires dans la législation ou la jurisprudence, cette question fera très probablement l'objet d'une analyse au cas par cas du système d'entraînement du modèle d'IA, afin d'évaluer si ses produits sont des élaborations proches des formes d'expression des œuvres initiales utilisées pour l'entraînement et/ou si les modèles utilisés par le système d'IA pour générer de nouvelles œuvres reproduisent des produits à peine discernables des œuvres originales de l'auteur.

3.2. Exception pour la fouille de textes et de données pour les données d'apprentissage

3.2.1. Applicabilité de l'exemption au titre de la fouille de textes et de données pour les données d'apprentissage de l'IA

L'IA en général et, surtout, les systèmes d'IA générative nécessitent de grands ensembles de données pour l'entraînement des machines et l'apprentissage profond¹⁰⁴, y compris des œuvres protégées par le droit d'auteur telles que de la musique, des images ou du texte, en fonction du résultat recherché. Ces besoins sont généralement satisfaits par la fouille de textes et de données (*text and data mining*, TDM), définie comme le processus automatisé d'extraction d'informations et de connaissances à partir de grandes quantités de textes et de données¹⁰⁵. Deux types de données peuvent être traités par TDM : tandis que la fouille de données traite les données structurées provenant de systèmes tels que les bases de données, les feuilles de calcul, etc., la fouille de textes traite les données non structurées trouvées dans les documents, les courriels, les médias sociaux et le web, où les modèles sont extraits de textes en langage naturel plutôt que de bases de données structurées et factuelles¹⁰⁶. La fouille de textes bénéficie des progrès réalisés dans le

¹⁰³ Voir E. Rosati, « [When is a derivative work original and thus protectable by copyright ? Classicist's critical edition makes its way to Luxembourg in fresh Romanian CJEU referral](#) », *The IPKat*. Voir aussi l'arrêt de la CJUE (cinquième chambre) du 11 juin 2020 dans l'affaire [C-833/18](#).

¹⁰⁴ Pour une distinction entre intelligence artificielle, apprentissage profond et apprentissage automatique, voir la vidéo

« [AI vs Machine learning vs. deep learning : know the differences](#) », *simpl;learn*,

¹⁰⁵ Une vue d'ensemble schématique des processus impliqués dans la fouille de textes de contenus académiques peut être consultée sur <https://libereurope.eu/topic/text-data-mining/>. Voir S. Ecolani, « Text and data mining : the copyright connection », in Campus G, Franzosi M. Pollicino O. *Digital Single Market and Artificial Intelligence*, Aracne éd., 2024, p. 799 et suiv.

¹⁰⁶ Hearst, M.A. « Text Data Mining », Mitkov, R. (éd.), *The Oxford Handbook of Computational Linguistics*, Oxford University Press : Oxford, UK, 2005, p. 616–662.



domaine du traitement du langage naturel, en particulier lorsqu'il s'agit de transformer un texte non structuré en données structurées pouvant faire l'objet d'une analyse.

Les activités de TDM deviennent critiques lorsqu'elles impliquent l'accès à des données et l'extraction à partir de contenus protégés par le droit d'auteur, ces activités pouvant potentiellement enfreindre les droits exclusifs reconnus par les lois nationales et les traités internationaux dont jouissent les auteurs et les titulaires de droits connexes, essentiellement des droits de reproduction et d'adaptation. Les activités de TDM importent également parce qu'elles sont au cœur de l'équilibre entre les droits des titulaires de droits et les droits des innovateurs, qui ont besoin d'une grande quantité de données pour développer des technologies susceptibles de favoriser l'innovation.

La règle fondamentale destinée à assurer cet équilibre conformément aux principes du cadre juridique international est le « test en trois étapes¹⁰⁷ », qui précise les critères que les États doivent prendre en compte lorsqu'ils introduisent des exceptions et des limitations aux droits exclusifs. Le test en trois étapes figure non seulement dans la Convention de Berne (article 9, paragraphe 2), mais aussi dans l'Accord sur les aspects des droits de propriété intellectuelle qui touchent au commerce (ADPIC, article 13¹⁰⁸), dans le Traité de l'OMPI sur le droit d'auteur (WCT, article 10¹⁰⁹) et dans le Traité de l'OMPI sur les interprétations et exécutions et les phonogrammes (WPPT, article 16¹¹⁰). Dans l'UE, le test en trois étapes est inscrit à l'article 5.5 de la Directive Infosoc¹¹¹, ainsi que dans d'autres directives

Dans d'autres juridictions, par exemple aux États-Unis, une approche différente a été adoptée, avec des exceptions potentiellement plus larges – à considérer à la lumière du test en trois étapes – en vertu du principe de l'usage loyal¹¹². Celui-ci permet d'évaluer

¹⁰⁷ Le test en trois étapes se trouve spécifiquement dans l'article 9(2) de la Convention de Berne et dans l'article 13 de l'Accord sur les ADPIC. Il stipule que toute limitation ou exception au droit d'auteur doit satisfaire à trois critères :

- a. Cas spéciaux : la limitation ou l'exception doit s'appliquer à certains cas spéciaux qui n'entrent pas en conflit avec l'exploitation normale de l'œuvre ;
- b. Absence de conflit : les exceptions ne doivent pas porter atteinte à l'exploitation normale de l'œuvre ; et
- c. Pas de préjudice déraisonnable : la limitation ou l'exception ne doit pas causer de préjudice injustifié aux intérêts légitimes des titulaires de droits.

La loi italienne sur le droit d'auteur 633 du 22 avril 1941 reprend littéralement les trois critères aux articles 69 bis, paragraphe 5, 70 *sexies*, 71 *bis*, paragraphe 3 *octies*, 71 *sexies*, paragraphe 4 et 71 *nonies*.

¹⁰⁸ [L'accord sur les ADPIC](#) est un protocole du GATT de l'Organisation mondiale du commerce. Il prévoit que les membres de l'OMC doivent se conformer aux dispositions de droit matériel de la Convention de Berne, à l'exception des dispositions relatives au droit moral des auteurs. Les accords internationaux conclus par l'Union font, dès leur entrée en vigueur, partie intégrante de l'ordre juridique de l'Union européenne (arrêts du 30 avril 1974, *Haegeman* (181/73, EU:C:1974:41, points 2/6) ; du 30 septembre 1987, *Demirel* (12/86, EU:C:1987:400, point 7) ; et du 8 mars 2011, *Lesoochranárske zoskupenie* (C 240/09, EU:C:2011:125, point 30). Ils lient donc les institutions de l'Union et ses États membres en vertu de l'article 216, paragraphe 2, du TFUE.

¹⁰⁹ <https://www.wipo.int/wipolex/fr/text/295168>.

¹¹⁰ <https://www.wipo.int/wipolex/fr/text/295580>.

¹¹¹ <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32001L0029>.

¹¹² Selon le Bureau américain du droit d'auteur (US Copyright Office), « l'usage loyal est une doctrine juridique qui promeut la liberté d'expression en autorisant l'utilisation sans licence d'œuvres protégées par le droit d'auteur dans certaines circonstances. L'article 107 de la loi sur le droit d'auteur (*Copyright Act*) définit le cadre légal permettant de déterminer s'il s'agit d'une utilisation équitable ». Cet article prévoit la prise en



au cas par cas si certaines utilisations d'œuvres protégées par le droit d'auteur sont admissibles à des fins transformatives et non commerciales.

3.2.2. La fouille de textes et de données et l'impact sur les droits de reproduction et d'extraction

La directive (UE) 2019/790 sur le droit d'auteur dans le marché unique numérique¹¹³ consacre ses articles 3 et 4 à la fouille de textes et de données, c'est-à-dire à l'utilisation de techniques analytiques automatisées pour analyser de grandes quantités de textes et de données à des fins de recherche, d'innovation et autres, dans le but de générer de nouvelles perspectives, de nouvelles connaissances et potentiellement de nouveaux produits, possiblement à partir de l'analyse de contenus protégés par le droit d'auteur. Compte tenu de l'essor de l'IA générative depuis novembre 2022 (avec le lancement de ChatGPT), il convient de souligner que, au moment où le législateur européen avait introduit l'exception au titre du TDM, le contexte technologique n'était pas centré sur la possibilité de générer de nouveaux contenus par le biais de l'IA à partir de données d'apprentissage potentiellement collectées sur la base de l'exception relative à la fouille de textes et de données.

L'article 3 introduit une exception plus large pour le TDM en faveur des institutions culturelles et de recherche, tandis que l'article 4 établit des conditions plus strictes pour l'exception générale au titre du TDM, notamment à des fins potentiellement commerciales¹¹⁴. Il existe toutefois certains aspects communs, tels que l'exemption des droits exclusifs en matière de reproduction et d'extraction. En ce qui concerne le droit de reproduction, les contenus protégés par le droit d'auteur sont potentiellement copiés sur les installations de stockage du responsable de la fouille des données et, grâce à la sélection automatique qui s'ensuit, ils sont copiés (et/ou adaptés) dans un nouvel ensemble de données au moyen du logiciel d'analyse ; cette reproduction peut être simplement transitoire et ne consister qu'en des fragments d'œuvres¹¹⁵. Pour ces

compte des quatre facteurs suivants dans l'évaluation d'une question d'usage loyal : 1. Le but et le caractère de l'usage, y compris sa nature commerciale ou à des fins éducatives non lucratives ; 2. la nature de l'œuvre protégée par le droit d'auteur ; 3. le volume et l'importance de la partie utilisée par rapport à l'œuvre protégée par le droit d'auteur dans son ensemble ; 4. l'incidence de l'usage sur le marché potentiel de l'œuvre protégée ou sur sa valeur.

¹¹³ Margoni T., Kretschmer M., 2018/04/25, [The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs](#).

¹¹⁴ Geiger C., Frosio G., Bulayenko O., « [The exception for Text and Data Mining \(TDM\) in the Proposed Directive on Copyright in the Digital Single Market Legal Aspects](#) », in *Centre for International Intellectual Property Studies* (CEIPI), article n° 2018-02,

¹¹⁵ Selon l'arrêt de la CJUE du 4 octobre 2011 dans les [affaires jointes C-403/08 et C-429/08 \(Premier League\)](#), §159, les reproductions partielles sont couvertes par le droit de reproduction de l'article 2 de la Directive Infosoc, lorsque les fragments « contiennent des éléments qui sont l'expression de la création intellectuelle propre aux auteurs, l'ensemble composé des fragments reproduits simultanément devant être examiné en vue de vérifier s'il contient de tels éléments ».



fragments, en l'absence d'exemption au droit d'auteur, le TDM nécessiterait l'autorisation des titulaires de droits.

Le terme « extraction » dans les dispositions relatives au TDM semble être une référence claire à l'exemption du TDM du droit *sui generis* qui réserve au fabricant « l'extraction ou la réutilisation d'une partie substantielle » de la base de données. Aucune référence explicite n'est faite à l'articulation entre le TDM et les droits d'adaptation ou de modification, qui peuvent être considérés comme limités au regard de l'article 12 de la Convention de Berne¹¹⁶, et qui représenteraient à coup sûr l'aspect central de la prise en compte de l'exception relative au TDM pour justifier l'entraînement des systèmes d'IA à partir de contenus protégés par le droit d'auteur.

Afin d'encourager l'innovation par le biais de l'exception en faveur du TDM également à des fins commerciales, l'article 4 introduit une exception générale pour les particuliers ou les organisations qui réalisent des activités de TDM. Entre le droit d'auteur, d'une part, et l'innovation et la recherche, d'autre part, il est plus complexe de parvenir à un juste équilibre que dans le cas de l'article 3, qui prévoit la possibilité d'accorder des licences pour l'utilisation de contenus protégés par le droit d'auteur à des fins de TDM. L'article 4 a défini cet équilibre au moyen du droit de « réserve », la prérogative que les titulaires de droits peuvent exercer au moyen d'une réserve exprimée « d'une manière appropriée ». Lorsque les contenus protégés par le droit d'auteur sont mis à disposition en ligne, la réserve doit être exercée par des moyens lisibles par machine.

À l'heure actuelle, quelques licences ont été annoncées entre les ayants droit et les plateformes (entre OpenAI et Associated Press¹¹⁷), tandis que le *New York Times* interdit l'utilisation de son contenu pour entraîner des modèles d'IA¹¹⁸ et que des médias français tels que Radio France et France 24 mettent en place des outils de protection contre le moissonnage de données¹¹⁹.

¹¹⁶ Article 12 – Droit d'adaptation, d'arrangement et d'autres transformations – « Les auteurs d'œuvres littéraires ou artistiques jouissent du droit exclusif d'autoriser les adaptations, arrangements et autres transformations de leurs œuvres ».

¹¹⁷ <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a>.

¹¹⁸ <https://www.theverge.com/2023/8/14/23831109/the-new-york-times-ai-web-scraping-rules-terms-of-service>.

¹¹⁹ <https://www.lesechos.fr/tech-medias/medias/ia-les-medias-francais-sorganisent-face-a-la-collecte-de-donnees-par-les-robots-1973079>



3.3. Textes législatifs relatifs à l'IA

3.3.1. Règlement de l'UE sur l'IA et droit d'auteur : règles de transparence et mesures relatives à la fouille de textes et de données

Le règlement établissant des règles harmonisées en matière d'intelligence artificielle (le « Règlement sur l'IA ») s'inscrit dans le cadre d'un projet beaucoup plus vaste et ambitieux mené par la Commission von der Leyen depuis 2019 déjà, qui comprend notamment le Livre blanc sur l'IA – Une approche européenne axée sur l'excellence et la confiance¹²⁰, ainsi que la proposition de directive relative à l'adaptation des règles de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle¹²¹. Parallèlement, le Parlement européen a également entrepris des efforts considérables dans le domaine de l'IA, en particulier en ce qui concerne des questions telles que l'éthique, la responsabilité et le droit d'auteur¹²², confirmant l'intention de l'UE de jouer un rôle de premier plan dans l'identification et la réglementation des questions de gestion et d'encadrement juridique de l'intelligence artificielle pour l'avenir.

Le choix d'un règlement – et donc de l'applicabilité directe dans les États membres de l'UE, conformément à l'article 288 du TFUE – plutôt que d'une directive est une indication claire de la direction prise par l'UE. Grâce au Règlement sur l'IA, l'UE sera effectivement en mesure d'injecter une discipline uniforme directement dans les cadres juridiques respectifs de chaque État membre, en principe sans nécessiter de transposition ou de mise en œuvre au niveau national.

Le 12 juillet 2024, le Règlement (UE) 2024/1689 du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle¹²³ (le Règlement sur l'IA) a été publié au Journal officiel de l'Union européenne. Il entrera en vigueur 20 jours après sa publication.

L'article 53 sur les « obligations pour les fournisseurs de modèles d'IA à usage général » a été introduit, avec deux exigences distinctes liées au droit d'auteur : (i) Le paragraphe 1(c) exige des fournisseurs de modèles d'IA à usage général qu'ils « mettent en place une politique visant à se conformer au droit de l'Union en matière de droit d'auteur et droits voisins, et notamment à identifier et à respecter, y compris au moyen de technologies de pointe, une réservation de droits exprimée conformément à l'article 4,

¹²⁰ Commission européenne, *Livre blanc sur l'IA – Une approche européenne axée sur l'excellence et la confiance*, COM(2020) 65 final, 2020.

¹²¹ Commission européenne, *Proposition de directive relative à l'adaptation des règles de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle* (Directive sur la responsabilité en matière d'IA), COM(2022) 496 final, 2022.

¹²² Parlement européen, *Résolution du 20 octobre 2020 concernant un cadre pour les aspects éthiques de l'intelligence artificielle, de la robotique et des technologies connexes*, 2020/2012(INL).

¹²³ Disponible sur https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=OJ:L_202401689.



paragraphe 3, de la directive (UE) 2019/790 » et (ii) le paragraphe 1(d) les oblige à élaborer et mettre « à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner le modèle d'IA à usage général, conformément à un modèle fourni par le Bureau de l'IA¹²⁴ ».

Le contenu du « résumé suffisamment détaillé » sera déterminé par un modèle qui sera élaboré par le Bureau de l'IA de l'UE. Le considérant 107 indique que le résumé doit être généralement complet plutôt que détaillé sur le plan technique, et indiquer par exemple « les principaux jeux ou collections de données utilisés pour entraîner le modèle ». Avant que le modèle ne soit disponible, les opérateurs devront développer de bonnes pratiques sectorielles¹²⁵.

D'autre part, la réservation des droits dans le cadre de l'exception relative au TDM – en particulier dans un environnement en ligne – pourrait, pour être efficace, impliquer le développement de « technologies de pointe » adéquates, qui sont probablement prises en compte dans la demande de normalisation déjà soumise par la Commission européenne aux organismes européens de normalisation (OEN¹²⁶).

3.3.2. IA et exception relative à la fouille de textes et de données : les propositions de lois de l'Italie et de la Pologne

Le 23 avril 2024, le **Gouvernement italien** a publié le texte d'un projet de loi¹²⁷ introduisant des dispositions réglementaires concernant l'utilisation de systèmes d'intelligence artificielle dans le système juridique italien (« Proposition de loi sur l'IA¹²⁸ »). Le texte a été approuvé par le Conseil des ministres, puis soumis au Parlement italien pour discussion le 20 mai 2024¹²⁹. En ce qui concerne les données d'apprentissage, l'article 24 de la proposition de loi sur l'IA introduit également un nouvel article 70 *septies* dans la loi italienne sur le droit d'auteur (« La reproduction et l'extraction d'œuvres ou d'autres matériaux par le biais de modèles et de systèmes d'intelligence artificielle, y compris génératifs, sont autorisées dans le respect des articles 70 *ter* et 70 *quater* »). L'article proposé semble destiné à renforcer le principe selon lequel, sauf dans le cas de la recherche scientifique, les titulaires de droits d'auteur peuvent s'opposer à l'utilisation de leur contenu dans le cadre de la fouille de textes et de données à des fins commerciales. Cette disposition est cohérente avec le principe déjà exprimé dans l'article 53, paragraphe 1(c) du règlement de l'UE sur l'IA.

¹²⁴ Keller P., [A first look at the copyright relevant parts in the final AI Act compromise](#).

¹²⁵ Voir Frank C. et Schmid G., [AI, the Artificial Intelligence Act & Copyright](#).

¹²⁶ Voir <https://artificialintelligenceact.eu/standard-setting/> et <https://www.etsi.org/newsroom/blogs/entry/standardization-request-in-support-of-safe-trustworthy-artificial-intelligence>.

¹²⁷ Voir Campus G., [Artificial Intelligence and copyright : the Italian AI Law Proposal](#).

¹²⁸ Voir <https://www.governo.it/it/articolo/comunicato-stampa-del-consiglio-dei-ministri-n-78/25501>.

¹²⁹ Voir <https://www.senato.it/service/PDF/PDFServer/DF/437373.pdf>.



La **Pologne** n'a pas encore finalisé la transposition en droit national des dispositions de la directive de 2019 sur le droit d'auteur dans le marché unique numérique. Le Gouvernement polonais affirme à propos de ce délai qu'il lui a permis d'examiner correctement l'impact de l'IA générative sur le droit d'auteur et d'arriver à la conclusion que l'entraînement de systèmes d'IA générative sur des œuvres protégées par le droit d'auteur n'entraîne pas dans le champ d'application des exceptions relatives à la fouille de textes et de données prévues par la directive, étant donné que ce type d'utilisation autorisée n'avait pas été conçu dans le contexte de l'intelligence artificielle¹³⁰.

3.4. L'impact de la jurisprudence

3.4.1. Vue d'ensemble des affaires relatives aux données d'apprentissage (États-Unis et Europe)

L'action collective sur le droit d'auteur contre OpenAI aux États-Unis : cette action collective a été déposée le 28 juin 2023¹³¹, devant la Cour de district des États-Unis – District Nord de Californie, Division de San Francisco, par deux auteurs (Paul Tremblay et Mona Awad), en leur nom propre et en celui des autres parties à l'action collective, à l'encontre d'OpenAI Inc. et autres. Les plaignants ont demandé un procès devant jury pour obtenir une injonction et des dommages-intérêts en conséquence du comportement prétendument illégal des défendeurs. Selon les plaignants, « *le produit d'un grand modèle de langage dépend [...] entièrement et uniquement du matériel contenu dans ses données d'apprentissage* » (voir le point I.3). Une grande partie du matériel contenu dans les données d'apprentissage d'OpenAI provient d'œuvres protégées par le droit d'auteur – y compris des livres écrits par les plaignants – qui ont été copiés par OpenAI sans consentement, sans citation et sans compensation. OpenAI n'a jamais révélé quels livres faisaient partie de ses ensembles de données Books1 et Books2, qui sont les « ensembles de données d'apprentissage [qui] proviennent de deux corpus de livres issus d'internet » (voir § V.30). OpenAI a justifié le manque d'informations sur la provenance des données par « le paysage concurrentiel et les implications en termes de sécurité des modèles à grande échelle » (voir § V.35).

L'action collective américaine contre Google Bard pour moissonnage de sites web : une autre action collective¹³² a été intentée contre Google devant la Cour de district des États-Unis – District nord de Californie, pour moissonnage de sites web présumé (couvrant à la fois des aspects liés au droit d'auteur et à la protection de la vie privée) dans le cadre de l'entraînement de ses outils d'IA, Bard, Imagen, MusicLM, Duet AI, et

¹³⁰ Voir Keller P., [TDM : Poland challenges the rule of EU copyright law](#).

¹³¹ Voir *Tremblay P. et Awad M. c. OpenAI INC. et al*, n° 3:23-cv-03223.

¹³² Voir *J. L. c. Alphabet Inc*, Cour de district des États-Unis – District nord de Californie, n° 3:23-cv-03440.



Gemini¹³³. Pour développer ses produits, le modèle d'IA de Google a été pré-entraîné sur environ 1,56 trillion de mots issus de « données de dialogues publics et de textes web », tirés d'Infiniset, un amalgame de contenus internet méticuleusement sélectionnés pour améliorer les capacités conversationnelles du modèle (§ 1.76¹³⁴). En outre, l'origine des données utilisées pour entraîner LaMDA¹³⁵, le modèle linguistique sur lequel repose Google Bard, comprend l'ensemble de données C4. La série de données C4, créée par Google en 2020, est tirée de l'ensemble de données Common Crawl, qui est un jeu de données disponible en libre accès, mais qui est destiné à des fins de recherche et d'éducation et qui, selon les plaignants, n'a jamais été prévu pour être transformé en un produit d'IA à des fins commerciales¹³⁶.

L'action collective américaine contre Meta LLaMA : deux actions collectives contre Meta sont soutenues par certains titulaires de droits d'auteur (principalement des auteurs de livres), à propos d'une prétendue violation de la propriété intellectuelle relative à leurs livres et œuvres écrites s'agissant du matériel de formation pour LLaMA (Large Language Model Meta AI). Cette jurisprudence est intéressante au regard de la reconstruction de la technologie déployée par Meta et de la méthodologie d'apprentissage (du moins du point de vue du plaignant), mais aussi parce que le tribunal a eu l'occasion d'évaluer de manière préliminaire la validité des revendications¹³⁷. La première action collective *Kadrey c. Meta* a été déposée le 7 juillet 2023¹³⁸ devant la Cour de district des États-Unis – District nord de Californie, Division de San Francisco. La seconde action collective *Chabon c. Meta* a été déposée le 12 septembre 2023 devant le même tribunal¹³⁹. Les deux plaintes reposent essentiellement sur les mêmes arguments et allégations factuelles. Meta note que 85 gigaoctets des données d'apprentissage proviennent d'une catégorie appelée « Books ». Selon les plaignants, cette catégorie comprend Bibliotik, une « bibliothèque fantôme » qui intéresse depuis longtemps les spécialistes de l'apprentissage de l'IA en raison de la grande quantité de documents protégés par le droit d'auteur qu'elle contient (y compris des œuvres des plaignants).

Le tribunal régional de Hambourg dans l'affaire LAION : En ce qui concerne l'UE, une affaire allemande est en cours devant le tribunal régional de Hambourg. Un photographe de stock poursuit l'organisation à but non lucratif LAION, qui propose l'ensemble de données LAION-5B destiné à entraîner de larges modèles image-texte. L'action en justice porte sur la copie illégale et vise à faire retirer ces images des données d'apprentissage. LAION, pour sa part, s'appuie notamment sur l'exception générale relative au TDM prévue à l'article 4 de la directive sur le droit d'auteur dans le marché

¹³³ <https://fingfx.thomsonreuters.com/gfx/legaldocs/myvmodloqvr/GOOGLE%20AI%20LAWSUIT%20complaint.pdf>.

¹³⁴ Voir <https://medium.com/@taureanjoewhat-sites-were-used-for-training-google-bard-ai-1216600f452d> et <https://www.searchenginejournal.com/google-bard-training-data/478941/#close>.

¹³⁵ Voir <https://arxiv.org/pdf/2201.08239.pdf>.

¹³⁶ Voir <https://commoncrawl.org/> et <https://www.forbes.com/sites/kalevleataru/2017/09/28/common-crawl-and-unlocking-web-archives-for-research/?sh=1e3d3c233b83>.

¹³⁷ Disponible sur

<https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.62.0.pdf>.

¹³⁸ Disponible sur <https://www.courtlistener.com/docket/67569326/kadrey-v-meta-platforms-inc/>.

¹³⁹ Disponible sur <https://www.courtlistener.com/docket/67785353/chabon-v-meta-platforms-inc/>.



unique numérique, mais aussi sur l'exception au titre du TDM à des fins de recherche scientifique en vertu de l'article 3 de cette directive (en raison de sa nature non lucrative), qui ne prévoit pas de possibilité de « réserve¹⁴⁰ ». Certaines conclusions préliminaires de la phase d'audience, telles qu'elles ont été rapportées¹⁴¹, soulignent que la Cour a considéré que les images litigieuses étaient « légalement accessibles » sur le site de photos de stock et qu'en vertu de l'article 44b de la loi allemande sur le droit d'auteur, les copies au titre de l'exception TDM ne peuvent être effectuées que « dans le but de recueillir des informations, en particulier en ce qui concerne les modèles, les tendances et les corrélations » (et la Cour penchait vers l'acceptation d'une utilisation à des fins de collecte de corrélations). Un autre point important en jeu concerne la manière appropriée de faire part de sa réserve, puisque l'article 44b de la loi allemande sur le droit d'auteur exige que celle-ci soit communiquée – dans un environnement en ligne – dans un format lisible par machine (cela signifie qu'un refus en texte clair n'est pas suffisant en ligne ; il faut exprimer son refus via un fichier robots.txt).

3.5. Quelques observations (préliminaires) sur la jurisprudence

Les affaires susmentionnées en sont pour la plupart à un stade précoce. Elles semblent néanmoins pertinentes pour un certain nombre de raisons. Tout d'abord, parce que dans leurs reconstructions factuelles, il semble évident que le problème crucial concerne les données d'apprentissage pour les plus grands modèles linguistiques. Par conséquent, certaines des premières dispositions du Règlement sur l'IA consacrées aux données d'apprentissage sont spécifiquement des règles de transparence visant à faire la lumière sur le processus d'entraînement.

Le deuxième point pertinent concerne les arguments utilisés par les fournisseurs d'IA générative pour répondre aux allégations des plaignants. Ils tirent parti du fait que les plaignants n'ont pas été en mesure de démontrer comment, sur la base du fonctionnement des larges modèles linguistiques, les données d'apprentissage sont converties en sorties et si elles peuvent être considérées comme des œuvres dérivées (les allégations soulignent notamment la similarité entre les œuvres utilisées pour l'apprentissage et les sorties).

Dans certains cas, une défense fondée sur *l'usage loyal* a également été introduite. *L'usage loyal* est une exception à la loi sur le droit d'auteur conçue pour permettre une utilisation limitée de matériel protégé par le droit d'auteur sans autorisation à des fins telles que le commentaire, la critique, le reportage et les études scientifiques¹⁴². Mais le contre-argument est que la collecte et l'utilisation par les parties en cause de matériel

¹⁴⁰ Voir <https://ceplic.org/news/an-up-date-on-the-robert-kneschke-v-laion-e-v> et <https://www.heise.de/hintergrund/Stock-photographer-sues-AI-association-LAIION-The-crux-with-AI-training-data-8988690.html>.

¹⁴¹ Voir Brúß M. [ici](#) et Graef O.R. [ici](#).

¹⁴² Voir *McGucken c. Pub Ocean Limited*, 42 F.4th 1149 (9^e Cir. 2022).



protégé par le droit d'auteur, sans possibilité pour les détenteurs de droits d'auteur de faire état de leur réserve, dépasserait la portée de l'interprétation juridique de *l'usage loyal*, étant donné que la copie d'une œuvre entière va à son encontre¹⁴³.

Il sera intéressant de voir si les décisions des tribunaux américains et européens convergeront ou non sur la question des données d'apprentissage, en tenant compte du fait que les approches américaine comme européenne des exceptions au droit d'auteur devraient être interprétées conformément au test en trois étapes prévu par la Convention de Berne.

¹⁴³ Voir *VHT c. Zillow Group*, 918 F.3^d 723, 743 (9^e Cir. 2019) ; *Worldwide Church of God c. Phila. Church of God, Inc.*, 227 F.3^d 110, 1118 (9^e Cir. 2000).

PART III – Le statut juridique des prompts dans l'IA générative

Lorsque l'on parle de contenus assistés ou générés par l'IA, l'accent est souvent mis sur les violations potentielles des droits d'auteur, en particulier dans l'industrie audiovisuelle, où les titulaires de droits s'inquiètent de l'utilisation de leurs œuvres comme données d'apprentissage.

Mais qu'en est-il du prompt lui-même ? Un prompt peut-il être protégé par la loi¹⁴⁴ ? Peut-il être considéré comme un secret commercial si son utilisation s'avère très profitable à l'entreprise ? Puisque le prompt donne des instructions à l'IA générative et influence son produit, devrait-on lui accorder une importance du point de vue juridique ?

En ce qui concerne la violation du droit d'auteur, les prompts sont utilisés par des particuliers, ce qui rend difficile pour les titulaires de droits de détecter s'ils peuvent entraîner une infraction. La rédaction de prompts devrait-elle faire l'objet d'un encadrement ? Un tel contrôle pourrait potentiellement entraver la liberté des utilisateurs de s'exprimer par leur biais. Toutefois, lorsqu'il s'agit d'équilibrer les intérêts, est-il justifié de limiter la liberté d'une personne pour protéger quelque chose de plus important ? La protection du droit d'auteur pourrait-elle avoir un effet sur la liberté d'expression ?

Ces questions restent théoriques à l'heure où nous écrivons ces lignes, et il faudra voir comment les cadres relatifs à la protection des droits de l'homme relèveront ces défis.

¹⁴⁴ [Rethinking Copyright Law: The Case for Protecting AI-Generated Content and Rewarding Those Who Truly Know What They Want, Ziyong "Sean" Li, Benesch, 14 mai 2024](#)



4. Paternité des œuvres, responsabilité et transparence dans le contexte des contenus générés par l'IA

Malte Baumann et Jan Bernd Nordemann¹⁴⁵, avocats, cabinet NORDEMANN, Berlin ; Jan Bernd Nordemann est également professeur honoraire à l'université Humboldt de Berlin.

4.1. Paternité

4.1.1. Le créateur humain en tant qu'auteur

En vertu du droit de l'Union européenne, seule une création humaine peut bénéficier de la protection du droit d'auteur. Pour satisfaire à la définition d'une œuvre, celle-ci doit refléter la personnalité de son auteur en tant qu'expression de ses choix libres et créatifs¹⁴⁶. Ainsi, l'accent est mis sur le créateur humain et sur les actions qu'il a effectuées pour façonner l'œuvre. Les auteurs sont en mesure de donner à leurs œuvres une « touche personnelle » grâce à leurs choix propres et à l'usage qu'ils font de la liberté dont ils disposent¹⁴⁷. Un effet purement esthétique qui ne résulte pas d'un choix personnel et créatif n'est pas suffisant pour justifier une protection en tant qu'œuvre¹⁴⁸. En outre, la protection du droit d'auteur ne peut être accordée si la conception d'un produit est dictée par des considérations, des règles ou des contraintes techniques¹⁴⁹.

Cette approche anthropocentrique du droit de de l'UE se retrouve non seulement dans le critère d'originalité, mais aussi dans la durée de la protection, calquée sur la vie

¹⁴⁵ Jan Bernd Nordemann (associé) et Malte Baumann (collaborateur), avocats.

¹⁴⁶ [Cofemel](#), 12 septembre 2019, C-683/1 ; [Eva-Marie Painer](#), arrêt du 1^{er} décembre 2011, C-145/10.

¹⁴⁷ [Eva-Marie Painer](#), arrêt du 1^{er} décembre 2011, C-145/10.

¹⁴⁸ [Cofemel](#), arrêt du 12 septembre 2019, C-683/17.

¹⁴⁹ [Football Dataco](#), arrêt du 1^{er} mars 2012, C-604/10.



de l'auteur¹⁵⁰. Dans la Convention de Berne, le concept de droit moral souligne cette approche centrée sur l'humain¹⁵¹.

Les tribunaux des États membres de l'UE (par exemple en République Tchèque) ont déjà appliqué ce principe et estimé que seul un être humain peut être considéré comme un auteur, et non l'IA¹⁵². Certains pays (comme la France) ont commencé à élaborer des propositions législatives sur le sujet, qui prévoient que l'IA ne peut pas être elle-même auteur d'une œuvre¹⁵³. Dans d'autres pays (comme l'Espagne), les lois sur le droit d'auteur ne laissent déjà aucune place au doute en stipulant explicitement que seules les personnes physiques peuvent être considérées comme des auteurs¹⁵⁴.

Aux États-Unis, le principe selon lequel seuls les humains peuvent être auteurs était déjà établi dans la loi sur le droit d'auteur avant même la nouvelle ère de l'IA¹⁵⁵. En conséquence, le tribunal du district de Columbia a estimé en 2023 que le matériel dont l'expression est entièrement imputable à un système artificiel exploité par une machine ne bénéficie pas de la protection du droit d'auteur¹⁵⁶. Le Bureau du droit d'auteur des États-Unis (US Copyright Office) a maintenu ce principe et refuse d'accorder la protection du droit d'auteur à du matériel purement généré par l'IA¹⁵⁷. Seul le matériel résultant de la créativité humaine peut être protégé à ce titre¹⁵⁸. Les accords de la Writers Guild of America (WGA) suivent également cette approche¹⁵⁹.

La Chine a également adopté le principe selon lequel les modèles d'IA ne peuvent pas être considérés comme des auteurs¹⁶⁰. Une œuvre protégée par le droit d'auteur nécessite toujours un acte intellectuel de la part d'une personne.

Dans certains pays (tels que le Royaume-Uni et l'Irlande), la protection des contenus générés par ordinateur est expressément reconnue¹⁶¹. Toutefois, même dans ce

¹⁵⁰ Article 1 de la [directive 2006/116/CE](#).

¹⁵¹ Article 6bis de la [Convention de Berne](#) pour la protection des œuvres littéraires et artistiques du 9 septembre 1886, OMPI Lex n° TRT/BERNE/009 ; également Hugenholtz, P.B. et Quintais J.P., « [Copyright and Artificial Creation : Does EU Copyright Law Protect AI-Assisted Output ?](#) », *IIC - International Review of Intellectual Property and Competition Law* 52, 2021, p. 1190-1216.

¹⁵² Cerri A., « [Czech court finds that AI tool DALL-E cannot be the author of a copyright work](#) », *The IPKat*, 15 avril 2024.

¹⁵³ Dreyfus, « [Comprendre le droit d'auteur français à l'ère de l'IA : une analyse critique des derniers développements](#) », *Dreyfus*, 19 janvier 2024.

¹⁵⁴ Article 5, [Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia](#) (loi espagnole sur le droit d'auteur du 12 avril 1996).

¹⁵⁵ Cour d'appel des États-Unis pour le neuvième circuit, [arrêt du 23 avril 2018](#), n° 16-15469 [888 F.3d 418].

¹⁵⁶ Cour de district des États-Unis pour le district de Columbia, [arrêt du 18 August 2023](#), action civile n° 22-1564 (BAH) [2023 WL 5333236 (D.D.C. 18 août, 2023)]

¹⁵⁷ Bureau du droit d'auteur des États-Unis, [Copyright Registration Guidance : Works Containing Material Generated by Artificial Intelligence](#), 16 mars 2023.

¹⁵⁸ *Ibid.*

¹⁵⁹ Article 72 B, [Memorandum of Agreement for the 2023 WGA Theatrical and Television Basic Agreement](#) (protocole d'accord sur l'accord de base WGA pour le cinéma et la télévision) du 25 septembre 2023.

¹⁶⁰ Tribunal de l'internet de Pékin, [arrêt du 27 novembre 2023](#), (2023) Jing 0491 Min Chu n° 11279.

¹⁶¹ Royaume-Uni : Section 178, [Copyright, Designs and Patents Act 1988](#), 15 novembre 1988, Irlande : Section 21 (f), [Copyright and Related Rights Act](#), 2000, 10 juillet 2000.



cas, la qualité d'auteur est attribuée à la personne qui a créé les conditions nécessaires à la production du matériel¹⁶². En Ukraine, un droit *sui generis* sur le contenu généré par ordinateur a récemment été introduit, qui est explicitement conféré au titulaire des droits sur le programme d'ordinateur¹⁶³.

4.1.2. La création d'œuvres assistée par l'IA

En vertu de la législation de l'UE, les contenus générés par un ordinateur seul ne bénéficient pas d'une protection en tant qu'œuvre protégée par le droit d'auteur. Dans la pratique, cependant, il existe souvent une certaine forme d'influence humaine¹⁶⁴. Cette contribution humaine peut théoriquement servir de base à la protection du droit d'auteur, car l'utilisation d'outils techniques n'exclut pas la protection du droit d'auteur¹⁶⁵.

Quelle forme doit prendre la contribution humaine pour que le produit de l'IA puisse être attribué à une personne en tant qu'œuvre créée par elle ? Il existe trois phases possibles au cours desquelles une personne peut exercer une influence décisive sur la production d'un contenu par l'IA :

- lors de la création et de la configuration du modèle et du système d'IA (sélection des données d'apprentissage ; programmation du système d'IA et définition de son usage ; apprentissage ciblé du modèle) ;
- dans les indications données à l'IA par le biais de prompts ;
- dans la révision et le remaniement de l'ébauche produite par l'IA.

Cela correspond aux phases créatives définies par la CJUE dans le contexte de la photographie de portrait, qui offrent des espaces de liberté créative : les étapes préparatoires, l'exécution proprement dite et la révision ultérieure du résultat¹⁶⁶.

Selon la CJUE, même un degré très limité de créativité humaine est suffisant pour justifier une protection en tant qu'œuvre. L'élément déterminant est l'existence d'une quelconque liberté de choix individuel¹⁶⁷. Il n'est pas nécessaire que cette liberté soit particulièrement grande ni qu'elle soit utilisée de manière révolutionnaire : même un simple extrait de 11 mots d'un journal quotidien peut constituer une œuvre protégée¹⁶⁸, de même qu'une simple photographie de portrait¹⁶⁹. En revanche, la simple collecte

¹⁶² Royaume-Uni : Section 9 (3), *Copyright, Designs and Patents Act 1988* ; Irlande : Section 21 (f), *Copyright and Related Rights Act*, 2000.

¹⁶³ *Закон України № 2811-IX від 01.12.2022 Про авторське право і суміжні права* (Loi n° 2811-IX du 1^{er} décembre 2022 sur le droit d'auteur et les droits voisins), amendée par n° 2974-IX du 20 mars 2023.

¹⁶⁴ Milityna K., « *Human Creative Contribution to AI-Based Output - One Just Can't Get Enough* », *GRUR Int.*, 2023 p. 939-949.

¹⁶⁵ *Eva-Marie Painer*, arrêt du 1^{er} décembre 2011, C-145/10.

¹⁶⁶ *Eva-Marie Painer*, arrêt du 1^{er} décembre 2011, C-145/10 ; Hartmann C. *et al.*, *Trends and developments in artificial intelligence*, Office des publications de l'Union européenne, septembre 2020, p. 73.

¹⁶⁷ *Football Dataco*, arrêt du 1^{er} mars 2012, C-604/10.

¹⁶⁸ *Infopaq*, arrêt du 16 juillet 2009, C-5/08.

¹⁶⁹ *Eva-Marie Painer*, arrêt du 1^{er} décembre 2011, C-145/10.



d'informations factuelles ne suffit pas¹⁷⁰. La plupart des *prompts* répondent à ces exigences.

Toutefois, il importe non seulement que la marge de décision soit utilisée de manière créative, mais aussi que ces décisions personnelles soient reflétées dans le produit final de l'IA. L'expression spécifique doit refléter les choix créatifs libres de la personne¹⁷¹. L'intervention de l'IA ne doit donc pas éclipser complètement l'apport de l'être humain. Cela est conforme au principe général selon lequel une simple idée en tant que telle ne peut bénéficier de la protection du droit d'auteur, mais seulement son expression concrète¹⁷².

La CJUE elle-même ne s'est pas encore prononcée sur des affaires spécifiques à la création d'œuvres en lien avec l'IA, mais il existe des arrêts nationaux et des décisions des autorités publiques en la matière. Ceux-ci montrent que des réponses plus ou moins strictes peuvent être apportées à la question de savoir quelles exigences spécifiques doivent être appliquées à la contribution créative humaine. En fin de compte, malgré les efforts des juristes pour développer des méthodes d'évaluation généralisées¹⁷³, il convient de procéder à une évaluation au cas par cas en tenant compte des normes applicables au niveau national. L'outil d'IA utilisé et le degré d'automatisation joueront un rôle tout aussi important que l'étendue et la qualité de la contribution humaine concernée.

En France, les tribunaux ont jusqu'à présent appliqué un ensemble de critères plutôt généreux. En effet, les juridictions inférieures ont très tôt établi que les créations assistées par ordinateur pouvaient également être protégées, tandis que la cour d'appel de Bordeaux n'a exigé qu'un degré minimal d'originalité humaine¹⁷⁴.

En Chine, un tribunal a statué sur une affaire dans laquelle un utilisateur avait formulé des indications détaillées et ciblées au moyen de plus de 100 *prompts*¹⁷⁵. Cela a suffi pour que le tribunal affirme la protection du droit d'auteur. Selon le tribunal, plus les instructions données dans les *prompts* sont claires, plus le résultat est susceptible de refléter l'expression humaine personnelle.

Le Bureau du droit d'auteur des États-Unis, quant à lui, est plus strict et considère les *prompts* comme de simples instructions données à l'IA. Il estime que l'IA est responsable de l'expression en question et que cela n'est généralement pas suffisant pour

¹⁷⁰ [Funke Medien](#), arrêt du 29 juillet 2019, C-469/17.

¹⁷¹ [Cofemel](#), arrêt du 12 septembre 2019, C-683/17 ; Milityna K., « [Human Creative Contribution to AI-Based Output - One Just Can\('t\) Get Enough](#) », *GRUR Int.*, 2023, p. 939-949.

¹⁷² Article 9(2), [Accord sur les aspects des droits de propriété intellectuelle qui touchent au commerce](#) (ADPIC) du 15 avril 1994 ; Article 2, [Traité de l'OMPI sur le droit d'auteur](#) (WCT) du 20 décembre 1996 ; Article 1(2) de la [directive 2009/24/CE](#).

¹⁷³ Milityna K., « [Human Creative Contribution to AI-Based Output - One Just Can\('t\) Get Enough](#) », *GRUR Int.*, 2023, p. 939-949 ; Hugenholtz P.B. et Quintais J.P., « [Copyright and Artificial Creation : Does EU Copyright Law Protect AI-Assisted Output ?](#) », *The International Review of Intellectual Property and Competition Law* 52, 2021, p. 1190-1216.

¹⁷⁴ Hartmann C. *et al.*, *Trends and developments in artificial intelligence*, Office des publications de l'Union européenne, septembre 2020, p. 82.

¹⁷⁵ Tribunal de l'internet de Pékin, [arrêt du 27 novembre 2023](#), (2023) Jing 0491 Min Chu n° 11279.



obtenir une protection par le droit d'auteur¹⁷⁶. Toutefois, le Bureau souligne également qu'un collage artistique de contenus générés par l'IA ou la révision par un humain d'un contenu généré par l'IA peut justifier une protection.

4.1.3. Qui est l'auteur ?

Dans la plupart des cas, si l'on considère qu'il y a un auteur, il s'agit de l'utilisateur du produit de l'IA. C'est le cas, par exemple, lorsqu'un utilisateur modifie les résultats produits par l'IA de manière à ce qu'ils reflètent sa personnalité. Comme nous l'avons déjà mentionné, il est également concevable qu'un utilisateur fournisse des instructions si spécifiques que l'expression concrète du résultat reflète suffisamment ses choix créatifs.

Si les choix créatifs effectués par les développeurs lors de la conception d'un système d'IA peuvent également donner lieu à une protection par le droit d'auteur, cette protection portera normalement sur le code du logiciel. Les choix effectués au cours du développement du système d'IA n'ont généralement pas un lien suffisant avec l'expression concrète du produit. En effet, la plupart des systèmes d'IA générative sont précisément destinés à un large éventail d'utilisations et ne sont pas créés pour produire une œuvre particulière¹⁷⁷. Les développeurs du système d'IA créent l'outil, mais pas l'œuvre.

4.1.4. La protection par les droits voisins

Certains droits voisins ne se rattachent pas à une réalisation créative humaine, mais protègent des investissements ou des efforts économiques et organisationnels. Ainsi, le droit voisin du producteur de films est particulièrement pertinent en ce qui concerne les produits de l'IA dans le secteur audiovisuel.

La législation de l'UE a partiellement harmonisé le droit voisin du producteur de films par le biais de directives¹⁷⁸. Selon l'article 2(1)(c) de la directive sur la location, tant les œuvres cinématographiques que les simples séquences animées d'images qui ne sont pas considérées comme des œuvres entrent dans la définition d'un « film ». En dehors de l'UE, aucun traité international ne régit le droit voisin du producteur de films¹⁷⁹.

¹⁷⁶ Bureau du droit d'auteur des États-Unis, « [Copyright Registration Guidance : Works Containing Material Generated by Artificial Intelligence](#) », 16 mars 2023.

¹⁷⁷ Militsyna K., « [Human Creative Contribution to AI-Based Output - One Just Can't Get Enough](#) », *GRUR Int.*, 2023, p. 939-949.

¹⁷⁸ Article 3(1)(d) de la [directive 2006/115/CE](#), article 3(3) de la [directive 2006/116/CE](#), et articles 2(d) et 3(2)(c) de la [directive 2001/29/CE](#).

¹⁷⁹ Loef R. et Verwehen U., « [One more Night – Überlegungen zum abgeleiteten fremdenrechtlichen Filmherstellerschutz](#) », *Zeitschrift für Urheber- und Medienrecht*, 2007, p. 706-711.



La création d'une vidéo à l'aide d'outils d'IA peut nécessiter un effort économique et organisationnel suffisant pour donner naissance à un droit voisin¹⁸⁰. La protection accordée peut être justifiée par l'effort requis pour se procurer le logiciel et le matériel, l'intégration dans les processus de travail et les produits, la conception du contenu, l'obtention des droits nécessaires ainsi que le savoir-faire engagé pour l'élaboration des *prompts*. Comme les simples images animées sont couvertes par la définition européenne du film, même les vidéos entièrement générées par l'IA pourraient bénéficier d'une protection. Il importe donc peu de savoir si le film peut ou non être protégé en tant qu'œuvre selon les critères énoncés ci-dessus¹⁸¹. Le droit voisin découle du support physique sur lequel le film a été placé en premier lieu, quel que soit son contenu. Le titulaire du droit est la personne ou l'entité qui fournit l'effort organisationnel et économique nécessaire.

Le droit voisin de l'organisme de radiodiffusion¹⁸² protège également le matériel diffusé, quel qu'en soit le contenu. Ainsi, un organisme de radiodiffusion peut recevoir des droits sur un contenu audiovisuel d'IA qui ne remplit pas les critères requis pour être protégé en tant qu'œuvre¹⁸³.

4.2. La responsabilité sur les produits générés par l'IA

4.2.1. Quand faut-il considérer qu'une infraction a été commise ?

En principe, on peut supposer que les règles générales traditionnelles et établies du droit d'auteur doivent être appliquées pour déterminer si un contenu peut ou non être considéré comme étant à l'origine d'une violation du droit d'auteur¹⁸⁴. En ce sens, les

¹⁸⁰ Baumann M., « [Presseleistungsschutzrecht: Der Schlüssel zum Schutz KI-generierter Erzeugnisse?](#) », AfP – *Zeitschrift für Medien- und Kommunikationsrecht*, 2024, p. 193-197 ; Ebers M. *et al.*, « § 9 KI und Urheberrecht », *Künstliche Intelligenz und Robotik*, Ebers M. *et al.* (éd.), C.H. Beck, 2020, marg. n° 70 ; Becker M., « [Das Urheberrecht als Trostpreis für den Menschen ? Die überraschende Verteilung von Leistungsschutzrechten für KI-Erzeugnisse](#) », *GRUR*, 2024, p. 505-514.

¹⁸¹ Hartmann C. *et al.*, *Trends and developments in artificial intelligence*, Office des publications de l'Union européenne, septembre 2020, p. 91.

¹⁸² Voir l'article 13 de la [Convention de Rome](#) sur la protection des artistes interprètes ou exécutants, des producteurs de phonogrammes et des organismes de radiodiffusion du 26 octobre 1961 ; l'article 3(2)(d), de la directive 2001/29/CE ; l'article 14(3) de l'accord sur les ADPIC.

¹⁸³ Becker M., « [Das Urheberrecht als Trostpreis für den Menschen? Die überraschende Verteilung von Leistungsschutzrechten für KI-Erzeugnisse](#) », *GRUR*, 2024, p. 505-514.

¹⁸⁴ Finke M., « [Urheberrechtliche Zulässigkeit der Nutzung des Outputs einer Künstlichen Intelligenz](#) », *Zeitschrift für Geistiges Eigentum*, 2023, p. 414-444, « [Generative KI : Eine "Blackbox" urheberrechtlicher Haftungsrisiken ? Balanceakt zwischen Innovationsförderung und effektivem Rechtsschutz für Werke Dritter](#) », *MMR – Zeitschrift für IT-Recht und Recht der Digitalisierung*, 2024, p. 298-304.



dispositions existantes de la législation de l'UE sur le droit d'auteur s'appliquent également à la production audiovisuelle générée par l'IA.

La législation européenne sur le droit d'auteur ne contient des dispositions explicites sur les adaptations que pour certains types d'œuvres, comme les logiciels¹⁸⁵ ou les bases de données protégées par le droit d'auteur¹⁸⁶. Pour les autres types d'œuvres, et en particulier pour les œuvres audiovisuelles, seules les lois sur le droit d'auteur des différents États membres de l'UE (par exemple en Belgique¹⁸⁷, en France¹⁸⁸ ou en Allemagne¹⁸⁹) contiennent des dispositions explicites sur les adaptations nécessitant une autorisation. Les droits de reproduction sont pleinement harmonisés dans la législation européenne en vertu de l'article 2 de la Directive Infosoc¹⁹⁰. L'harmonisation couvre également la reproduction partielle¹⁹¹. En ce qui concerne le droit de reproduction en lien avec le droit voisin du producteur de phonogrammes, la CJUE a estimé que les reproductions modifiées relevaient également de la notion de reproduction si l'original restait reconnaissable malgré la modification¹⁹².

Selon le Bundesgerichtshof (Cour fédérale de justice allemande), cette jurisprudence est également applicable au droit de reproduction d'un droit d'auteur au sens de l'article 2(a), de la Directive Infosoc¹⁹³. Toutefois, la Svea hovrätt, Patent- och marknadsöverdomstolen (cour d'appel de Svea, cour d'appel des brevets et du marché) a posé une question à la CJUE pour obtenir des éclaircissements sur l'applicabilité de la jurisprudence Pelham de la CJUE sur le droit voisin du producteur de phonogrammes aux droits d'exploitation du titulaire du droit d'auteur¹⁹⁴. On ne sait pas exactement comment la CJUE se prononcera sur cette question. Mais si la CJUE applique sa jurisprudence dans l'affaire Pelham au droit de reproduction dans le cadre du droit d'auteur en tant que tel, conformément à l'article 2 de la Directive Infosoc, le seul facteur pertinent en vertu du droit de l'UE pour déterminer s'il y a eu infraction est la possibilité de reconnaître l'œuvre. En conséquence, la question qui se pose serait alors la suivante : les éléments protégés par le droit d'auteur d'une œuvre antérieure peuvent-ils être reconnus dans l'œuvre nouvellement créée¹⁹⁵ (ultérieure) ?

Si l'on applique ce principe aux produits issus de l'IA, la question qui se pose est de savoir si des éléments protégés par le droit d'auteur provenant d'œuvres antérieures sont reconnaissables dans les contenus générés par l'IA. Cela dit, il n'existe pas de principe de priorité dans la législation sur le droit d'auteur, ce qui signifie que les

¹⁸⁵ Article 4(1)(b) de la directive 2009/24/CE.

¹⁸⁶ Article 5(b) de la [directive 96/9/CE](#).

¹⁸⁷ Article 1(1), [loi n° 2006-961 du 1^{er} août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information](#).

¹⁸⁸ L122-4, [Code de la propriété intellectuelle](#).

¹⁸⁹ § 23 (1), [Urheberrechtsgesetz](#) (loi allemande sur le droit d'auteur).

¹⁹⁰ [Infopaq](#), arrêt du 16 juillet 2009, C-5/08.

¹⁹¹ [Infopaq](#), arrêt du 16 juillet 2009, C-5/08.

¹⁹² [Pelham](#), arrêt du 29 juillet 2019, C-476/17.

¹⁹³ Bundesgerichtshof, [arrêt du 7 avril 2022](#), I ZR 222/20.

¹⁹⁴ [Mio et autres](#), demande de décision préjudicielle, C-580/23.

¹⁹⁵ Bundesgerichtshof, [arrêt du 7 avril 2022](#), I ZR 222/20 ; [Mio et autres](#), Demande de décision préjudicielle, C-580/23.



créations indépendantes ne sont pas considérées comme contrevenant au droit d'auteur. Elles ne sont pas assimilées à des reproductions de l'œuvre antérieure, car l'auteur de l'œuvre ultérieure a créé son œuvre de manière indépendante. À ce jour, il n'existe pas de jurisprudence de la CJUE sur la définition de ce qui constitue une création indépendante. Néanmoins, le principe selon lequel les créations indépendantes sont autorisées est largement reconnu dans l'ensemble de l'UE¹⁹⁶. Historiquement, l'argument selon lequel une œuvre est une création indépendante n'a été retenu qu'en de rares occasions. En Allemagne, par exemple, les similitudes de parties créatives données suffisent généralement à prouver à première vue que l'œuvre en question est une reproduction modifiée nécessitant une autorisation. Cette présomption ne peut être réfutée que s'il est probable, d'après les circonstances de l'espèce, que les similitudes s'expliquent autrement que par le fait que le créateur de l'œuvre ultérieure se soit inspiré de l'œuvre antérieure¹⁹⁷.

En ce qui concerne la création indépendante, il convient de se demander si ce principe peut s'appliquer aux produits générés par l'IA. On peut supposer que oui¹⁹⁸. Le principe serait le suivant : si un système d'IA a été entraîné sur une œuvre reconnaissable, son produit constitue une violation du droit d'auteur, mais si un système d'IA génère un produit qui est similaire à une autre œuvre par pure coïncidence, sans que l'IA ait été entraînée sur cette œuvre, il s'agira d'une création indépendante acceptable¹⁹⁹. Les règles relatives à l'hypothèse *prima facie* doivent également être appliquées ici.

4.2.2. Exceptions et limitations au droit d'auteur applicables aux contenus générés par l'IA

En fonction de la production spécifique, des exceptions et des limitations au droit d'auteur peuvent s'appliquer, notamment celles prévues à l'article 5(3)(k) de la Directive Infosoc pour la parodie, la caricature ou le pastiche. D'autres exceptions et limitations, concernant spécifiquement le contenu audiovisuel, doivent également être prises en

¹⁹⁶ Voir Peukert A., « [Copyright in the Artificial Intelligence Act – A Primer](#) », *GRUR Int.*, 2024, p. 497-509 ; Iaia V., « [To Be or Not to Be... Original Under Copyright Law, That Is \(One of\) the Main Questions Concerning AI-Produced Works](#) », *GRUR Int.*, 2022, p. 807-812 ; Inguanez D., « [A Refined Approach to Originality in EU Copyright Law in Light of the ECJ's Recent Copyright/Design Cumulation Case Law](#) », *International Review of Intellectual Property and Competition Law*, 2020, p. 797-822 ; sur la situation juridique au Royaume-Uni : Guadamuz A., « [A Scanner Darkly : Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs](#) », *GRUR Int.*, 2024, p. 111-127.

¹⁹⁷ Bundesgerichtshof, [arrêt du 3 février 1988](#), I ZR 142/86.

¹⁹⁸ Sur ce point de vue, à propos de la législation britannique en matière de droits d'auteur : Guadamuz A., « [A Scanner Darkly : Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs](#) », *GRUR Int.*, 2024, p. 111-127 ; sur la législation allemande en matière de droit d'auteur : Käde A., « [Do You Remember ?](#) », *Zeitschrift für Urheber- und Medienrecht*, 2024, p. 174-183 ; Nordemann J.B., « [Generative AI, copyright infringements and liability – My guess for a hot topic in 2024](#) », *Kluwer Copyright Blog*, 23 janvier 2024 ; Baumann M., « [Generative KI und Urheberrecht - Urheber und Anwender im Spannungsfeld](#) », *Neue Juristische Wochenschrift*, 2023, p. 3673-3678.

¹⁹⁹ Baumann M., « [Generative KI und Urheberrecht - Urheber und Anwender im Spannungsfeld](#) », *Neue Juristische Wochenschrift*, 2023, p. 3673-3678.



compte. La disposition de l'article 4 de la directive sur le droit d'auteur et les droits voisins dans le marché unique numérique relative à la fouille de textes et de données ne s'applique toutefois pas aux contenus produits par l'IA.

4.2.3. Responsabilité de l'utilisateur

Il n'existe pas de dispositions spécifiques à l'IA au niveau de l'UE qui traitent de la question de la responsabilité. Néanmoins, il existe des arguments convaincants en faveur de l'application des principes existants, quoique légèrement modifiés, à la question de savoir qui est responsable en ce qui concerne l'utilisation des contenus générés par l'IA.

Tout d'abord, les règles générales devraient s'appliquer lorsque les utilisateurs de l'IA utilisent eux-mêmes des contenus générés par l'IA d'une manière qui a une incidence sur le droit d'auteur. Cela signifierait que les utilisateurs d'IA sont responsables chaque fois qu'ils reproduisent les éléments issus de l'IA (article 2 de la Directive Infosoc), qu'ils les distribuent (article 4 de la Directive Infosoc) ou qu'ils les communiquent au public (article 3 de la Directive Infosoc). Bien que l'utilisateur effectue directement l'utilisation de l'IA et soit responsable en tant qu'auteur, il est concevable qu'il agisse par négligence (absence de faute), ce qui exclut toute demande de dommages-intérêts. En fin de compte, toute personne utilisant des outils d'IA devra vérifier si leur résultat contient des éléments d'œuvres de tiers – même s'il n'est pas toujours possible d'avoir une certitude à cet égard. Si un contenu issu de l'IA portant atteinte à des droits est communiqué au public par l'intermédiaire d'un fournisseur d'hébergement, le modèle de responsabilité de la CJUE tiré de l'affaire YouTube et Cyando²⁰⁰ est applicable. Ce modèle de responsabilité considère que les plateformes vidéo et autres fournisseurs d'hébergement sont responsables des violations du droit de communication au public qu'ils ont indirectement causées, à condition qu'ils aient joué un rôle incontournable dans la violation et qu'ils aient manqué à leurs obligations de diligence.

Il est également possible d'envisager la responsabilité du fournisseur d'IA. Il n'existe actuellement aucune responsabilité spécifique du fournisseur pour les violations du droit d'auteur dans le domaine de l'IA générative. Le projet de révision de la Directive sur la responsabilité du fait des produits défectueux²⁰¹ de la Commission européenne inclut expressément les logiciels en tant que produits, conformément à la définition de l'article 4(1). Toutefois, les violations du droit d'auteur n'entrent toujours pas dans le champ des dommages couverts par cette directive²⁰². Le projet de directive sur la responsabilité en matière d'IA ne contient pas non plus de dispositions relatives à la détermination de la responsabilité en cas de violation du droit d'auteur par des contenus générés par l'IA, et se contente de faciliter la mise en œuvre des droits²⁰³.

²⁰⁰ [YouTube et Cyando](#), arrêt du 22 juin 2021, C-682/18 et C-683/18.

²⁰¹ [COM\(2022\) 495 final](#).

²⁰² Baumann M., « Generative KI und Urheberrecht - Urheber und Anwender im Spannungsfeld », *Neue Juristische Wochenschrift*, 2023, p. 3673-3678.

²⁰³ *Ibid.*



Les règles existantes doivent donc être appliquées aux fournisseurs d'IA. Il convient toutefois de noter que, dans la plupart des cas, cela ne concernera que la responsabilité en cas de reproduction non autorisée sous la forme de contenus tirés de l'IA (article 2 de la Directive Infosoc).

Une solution possible consiste à utiliser le modèle de responsabilité de la CJUE pour les parties qui causent indirectement des infractions, comme cela a déjà été mentionné ci-dessus. Jusqu'à présent, le système de responsabilité de la CJUE n'a été appliqué qu'au droit de communication au public. Il serait toutefois logique de l'étendre à d'autres droits d'exploitation tels que le droit de reproduction²⁰⁴. Un argument particulier en faveur de cette extension est qu'il pourrait être utilisé pour créer un système de responsabilité équilibré pour les fournisseurs d'IA également. En outre, ces contenus ne sont pas purement générés par les utilisateurs, comme c'est le cas, par exemple, sur les plateformes d'hébergement classiques. Le contenu est déterminé dans une large mesure par le système d'IA. Les fournisseurs d'IA disposent également de moyens (recours à des filtres, listes noires, « *metaprompts* », équipes rouges (« *red teaming* »), méthodes d'entraînement, etc.) pour intégrer dans le système lui-même des mesures visant à réduire le risque d'infractions. À cet égard, le Règlement sur l'IA dispose que les fournisseurs de modèles d'IA à usage général doivent élaborer une politique visant à se conformer au droit de l'Union en matière de droit d'auteur (article 53, paragraphe 1, point c)). Dans ce contexte, il est logique que la responsabilité des utilisateurs comme des fournisseurs soit engagée et que des obligations de diligence soient imposées aux uns et aux autres.

Pour que le modèle de responsabilité de la CJUE s'applique, il faut tout d'abord que le fournisseur d'IA joue un rôle incontournable. Dans l'affaire *Ocilion IPTV Technologies*²⁰⁵, la CJUE a estimé que les fournisseurs de matériel et de logiciels ne jouaient pas un rôle incontournable car ils n'étaient pas en mesure d'influencer l'infraction. Dans cette affaire, le fournisseur de logiciels et de matériel informatique proposait ses services à un tiers et seul ce dernier était en contact avec les clients finaux qui étaient les destinataires des infractions²⁰⁶. Il convient toutefois de noter que dans cette affaire, les utilisateurs finaux n'étaient pas non plus les auteurs de des infractions (au droit de communication au public) ; ils se contentaient d'en être les destinataires et ils n'exerçaient aucune influence sur les infractions elles-mêmes. Par conséquent, il est possible qu'un fournisseur de logiciels et de matériel – tel qu'un fournisseur d'IA – joue le rôle incontournable requis s'il met l'infraction (sous la forme d'une reproduction non autorisée) à la disposition de l'utilisateur. Un fournisseur d'IA qui transmet une infraction à un utilisateur serait donc réputé jouer un rôle incontournable. Le fournisseur d'IA serait alors tenu de respecter certaines obligations de diligence pour ne pas voir sa responsabilité engagée.

²⁰⁴ Nordemann J.B., « [Neu : Täterschaftliche Haftung von Host Providern im Urheberrecht bei \(Verkehrs\) Pflichtverletzungen im Internet](#) », *Zeitschrift für Urheber- und Medienrecht*, 2022, p. 806-816, ; Nordemann J.B., « [Generative AI, copyright infringements and liability – My guess for a hot topic in 2024](#) », *Kluwer Copyright Blog*, 23 janvier 2024 ; pour un autre point de vue : [Sony Computer Entertainment Europe](#), conclusions de l'avocat général Szpunar du 25 avril 2024, C-159/23.

²⁰⁵ *Ocilion IPTV Technologies*, arrêt du 13 juillet 2023, C-426/21.

²⁰⁶ *Ibid.*



Une évaluation mesurée et donc une formulation nuancée des obligations de diligence semblent s'imposer lorsqu'il s'agit de contenus générés par l'IA qui portent atteinte au droit d'auteur. Le système d'IA peut dans une large mesure déterminer le contenu généré. Par conséquent, la fourniture d'un système d'IA ne se limite pas à mettre à disposition un simple logiciel permettant aux utilisateurs de créer des reproductions à leur guise. Les obligations de diligence doivent être assignées en fonction de la personne qui a causé le contenu illicite proprement dit.

Dans la mesure où l'IA n'est qu'un outil technique pour l'utilisateur et puisque ses paramètres clés pour déterminer le contenu sont fixés par l'utilisateur de l'IA (par exemple par le biais de ses *prompts*), lui seul peut être considéré comme l'auteur de l'infraction. C'est le cas par exemple si les *prompts* de l'utilisateur de l'IA sont conçus pour générer un contenu illicite. La situation est toutefois différente si les infractions sont principalement imputables à l'IA génératrice. Dans ce cas, la responsabilité pourrait revenir au fournisseur de l'IA sur la base d'un manquement aux obligations de diligence. Ce serait le cas si, par exemple, un utilisateur n'avait saisi que des indications anodines dans ses *prompts* et que le système d'IA avait généré l'infraction de manière autonome. Le fournisseur d'IA devrait à tout le moins être tenu par une obligation de diligence visant à prévenir les violations manifestes du droit d'auteur. Cette obligation s'appliquerait même si le fournisseur d'IA ignorait jusque-là que son système était capable de générer l'infraction en question.

Si l'utilisateur est le principal responsable des produits générés par l'IA, alors les obligations de diligence doivent être limitées. Toutefois, même dans ce cas, le fournisseur de l'IA pourrait endosser une certaine responsabilité. Après tout, l'IA génère bel et bien le contenu et joue un rôle incontournable dans l'infraction (voir ci-dessus). Une façon possible de traiter ces cas pourrait être d'appliquer les trois obligations de diligence énoncées dans l'affaire YouTube et Cyando²⁰⁷ d'une manière légèrement modifiée²⁰⁸ :

- Dès qu'il a connaissance du problème, le fournisseur d'IA doit faire tout ce qui est possible sur le plan technologique et que l'on peut raisonnablement attendre de lui, afin d'empêcher que l'infraction ne se reproduise.
- S'il sait ou devrait savoir que des utilisateurs reproduisent illégalement des contenus protégés par l'intermédiaire de son système, le fournisseur d'IA doit mettre en place les mesures techniques appropriées qu'il est permis d'attendre d'un fournisseur normalement diligent dans sa situation, afin de contrer de manière crédible et efficace les violations du droit d'auteur.
- Le fournisseur d'IA ne saurait participer à la sélection de contenus protégés qui sont reproduits illégalement, ni fournir des outils spécifiquement destinés à la reproduction illégale de ces contenus, ni promouvoir sciemment ces reproductions. Un facteur qui pourrait suggérer qu'un fournisseur d'IA encourage sciemment une telle reproduction serait que le fournisseur ait adopté un modèle

²⁰⁷ YouTube et Cyando, arrêt du 22 Juin 2021, C-682/18 et C-683/18.

²⁰⁸ Nordemann, J.B., « Generative Künstliche Intelligenz : Urheberrechtsverletzungen und Haftung », *GRUR*, 2024, p. 1-2.



économique incitant ses utilisateurs à reproduire des contenus protégés sous forme de produits issus de l'IA.

Il conviendra d'examiner plus avant si cette voie peut être poursuivie.

4.2.4. Les conditions d'utilisation des fournisseurs d'IA

Les conditions d'utilisation de la plupart des fournisseurs d'IA interdisent aux utilisateurs de générer des contenus illégaux et prévoient que l'utilisateur est seul responsable des contenus générés. Ces dispositions sont souvent accompagnées de clauses d'indemnisation qui protègent le fournisseur d'IA²⁰⁹. Ainsi, les fournisseurs d'IA tentent de se décharger, dans la mesure du possible, de toute responsabilité pour les contenus.

Ces dispositions n'ont généralement aucun effet sur la détermination de la responsabilité, du point de vue des tiers tels que les titulaires de droits. Elles sont toutefois très pertinentes en ce qui concerne la relation contractuelle entre l'utilisateur et le fournisseur. Si un utilisateur génère (délibérément) du matériel illicite, il enfreint ses obligations contractuelles et le fournisseur d'IA a des droits de recours contre lui si le fournisseur lui-même fait l'objet de réclamations pour violation des droits.

Ainsi, le fait que certains fournisseurs d'IA proposent des clauses d'indemnisation pour les contenus portant atteinte au droit d'auteur est particulièrement important pour les utilisateurs²¹⁰. Toutefois, les fournisseurs d'IA y intègrent souvent un large catalogue d'exceptions qui limitent considérablement les possibilités d'indemnisation.

4.2.5. Réduire la responsabilité éventuelle

Aucune analyse concluante n'a encore été réalisée sur la probabilité que des systèmes d'IA individuels génèrent des contenus contraires au droit. Toutefois, il apparaît déjà que cette possibilité sera largement déterminée par l'objectif spécifique du système d'IA concerné, ses données d'apprentissage²¹¹ ou son mode d'entraînement ainsi que par les *prompts*²¹².

²⁰⁹ De telles clauses d'indemnisation figurent dans presque toutes les conditions d'utilisation des fournisseurs d'IA. Voir à titre d'exemple la section 11.3 des [conditions d'utilisation de l'IA Mistral](#).

²¹⁰ Voir, par exemple, le point 50.10 [des conditions d'utilisation d'AWS ou l'engagement en matière de droits d'auteur vis-à-vis du client pour les services OpenAI de Microsoft Azure](#).

²¹¹ Milityna K., « [Human Creative Contribution to AI-Based Output - One Just Can't Get Enough](#) », *GRUR Int.*, 2023, p. 939-949 ; Pesch P.J. et Böhme R., « [Artpocalypse now? - Generative KI und die Vervielfältigung von Trainingsbildern](#) », *GRUR*, 2023, p. 997-1007.

²¹² Marcus G. et Southen R., « [Generative AI Has a Visual Plagiarism Problem. Experiments with Midjourney and DALL-E 3 show a copyright minefield](#) », *IEEE Spectrum*, 24 juin 2024 ; Carlini N. *et al.*, « [Extracting Training Data from Diffusion Models](#) », *arXiv*, 30 janvier 2023 ; Henderson P. *et al.*, « [Foundation Models and Fair Use](#) », *Journal for Machine Learning Research* 24, 2023, p. 1-79.



Il existe d'autres moyens de minimiser le risque de générer des contenus portant atteinte au droit. L'un d'entre eux consiste à utiliser un outil d'IA qui n'a été entraîné qu'avec du matériel propre ou du matériel autorisé. Même si le modèle de base a été entraîné avec un large éventail d'œuvres, un deuxième entraînement spécifique utilisant du matériel propre pourrait réduire la tendance du système à générer des contenus violant les droits d'auteur.

Lorsqu'ils rédigent leurs *prompts*, les utilisateurs peuvent s'abstenir de faire référence à des œuvres protégées. Cette mesure de précaution a un effet équivalent à celui de la « liste noire » du fournisseur, qui bloque certains *prompts*. En outre, les *metaprompts* offrent la possibilité de rédiger des instructions générales pour le système. Ces *metaprompts* peuvent être utilisés pour réduire la probabilité que le système d'IA génère un contenu portant atteinte aux droits²¹³.

Les fournisseurs d'IA peuvent procéder à des évaluations régulières pour évaluer le risque systémique de violation des droits (recours à des équipes rouges ou « *red teaming* ») et supprimer de leurs résultats les contenus signalés comme portant atteinte aux droits²¹⁴.

4.2.6. Transparence

Le fait d'indiquer en toute transparence que le contenu en question a été généré artificiellement protège les destinataires, qu'il s'agisse de consommateurs, ou d'utilisateurs finaux agissant à titre professionnel. Pour ces derniers, il est particulièrement important que le contenu puisse bénéficier d'une protection afin qu'ils puissent le concéder sous licence. En outre, les clients doivent être en mesure de déterminer si leurs fournisseurs utilisent l'IA afin d'évaluer leur responsabilité potentielle. Les garanties habituelles dans le domaine du cinéma, selon lesquelles tous les droits sur le matériel fourni sont détenus par le fournisseur, doivent faire l'objet d'un examen minutieux. En cas d'utilisation intensive de l'IA, il n'est pas possible de donner une telle garantie de manière absolument certaine. Par conséquent, le nombre d'utilisateurs exigeant une divulgation détaillée du recours à l'IA dans leurs accords contractuels va croissant.

Il convient de mentionner qu'il pourrait être important pour l'utilisateur de divulguer l'utilisation de l'IA et d'en informer spécifiquement ses partenaires contractuels, ne serait-ce que pour des raisons de responsabilité. Cela s'applique en particulier si l'intention est d'accorder des droits d'exploitation exclusifs sur les contenus produits. En vertu des accords de la Writers' Guild aux États-Unis, les scénaristes doivent même

²¹³ Voir [Microsoft Azure](#), engagement avec le client en matière de droits d'auteur ; mesures de précaution requises.

²¹⁴ *Ibid.*



obtenir l'autorisation préalable de leurs clients (producteurs de films) s'ils veulent utiliser l'IA²¹⁵.

Le Règlement sur l'IA prévoit également certaines obligations de transparence. En vertu de son article 50, paragraphe 1, les fournisseurs de systèmes d'IA doivent veiller à ce que les utilisateurs soient informés du fait qu'ils interagissent avec l'IA. Toutefois, cette obligation ne s'applique qu'à l'interaction directe et non à toutes les formes de contenus générés par l'IA. En revanche, les fournisseurs de systèmes d'IA à usage général sont tenus d'étiqueter tous ses résultats. L'article 50, paragraphe 2, du Règlement sur l'IA dispose que ces fournisseurs doivent veiller à ce que toutes les sorties des systèmes d'IA soient identifiées, dans un format lisible par machine, comme ayant été générées artificiellement.

Les personnes qui utilisent des outils d'IA pour créer des hypertrucages ou *deepfakes* doivent le déclarer (article 50, paragraphe 4, du Règlement sur l'IA). Il convient de noter à cet égard que la définition de l'hypertrucage à l'article 3(60) du Règlement sur l'IA, est très large : elle couvre toute image ou contenu audio ou vidéo présentant une ressemblance avec des personnes, des objets ou des lieux existants et qui pourrait être perçu par une personne comme authentique. Le caractère dangereux de la supercherie pour les personnes concernées n'entre pas en ligne de compte. Dans le domaine du cinéma en particulier, l'IA peut être utilisée pour générer des lieux réels comme décors artificiels ou des sosies de personnes réelles. Il s'agit alors d'hypertrucages.

En conclusion, l'utilisation de l'IA affecte un certain nombre de domaines différents du droit d'auteur. Le Règlement sur l'IA régit la question sous l'angle de la sécurité des produits et laisse ouvertes de nombreuses questions concernant l'application des règles générales traditionnelles et établies du droit d'auteur. Malgré l'autonomie des systèmes d'IA, l'accent reste mis sur l'être humain : la protection du droit d'auteur dépend de la contribution de l'être humain et la responsabilité dépend du respect par l'être humain de ses obligations de diligence. Les obligations de transparence concernant les produits générés par l'IA visent également principalement les intérêts légitimes des destinataires humains.

²¹⁵ Article 72 D, protocole d'accord sur l'accord de base WGA pour le cinéma et la télévision du 25 septembre 2023.



5. Droits de la personnalité et transparence

Kelsey Farish, Avocate spécialisée dans les affaires commerciales des médias et du divertissement, Reviewed & Cleared, Londres

5.1. Le contexte

Lorsque l'entreprise d'intelligence artificielle OpenAI a lancé le système ChatGPT-4o en mai 2024²¹⁶, les impressionnantes capacités de synthèse vocale du modèle, notamment sa capacité à imiter de manière apparemment impeccable les intonations vocales de nombreuses langues, ont été saluées. Mais la star hollywoodienne Scarlett Johansson a remarqué autre chose : selon elle, la voix de la plateforme est si étrangement similaire à la sienne que même ses amis les plus proches ne pouvaient pas les distinguer²¹⁷.

Ce qui complique les choses, c'est que, selon Johansson, OpenAI l'aurait approchée à plusieurs reprises pour qu'elle prête officiellement sa voix au produit, ce qu'elle a refusé. Lorsqu'une voix semblable à celle de Scarlett a tout de même été utilisée, l'actrice a été « choquée, furieuse et incrédule », ce qui a conduit son équipe juridique à demander des détails sur la façon dont la voix de ChatGPT-4o avait été développée. « À une époque où nous sommes tous confrontés à des *deepfakes* et où il nous faut protéger notre propre image, notre travail, notre identité, je pense que ces questions méritent une clarté absolue », a-t-elle expliqué dans un communiqué de presse.

Bien qu'OpenAI ait rapidement désactivé la voix du système, cet incident met en évidence les difficultés liées à la protection de l'image d'une personne à l'ère des contenus issus de l'intelligence artificielle générative (« *GenAI* »), tels que les *deepfakes*. De plus en plus souvent appelées « doubles numériques », « répliques » ou « clones », les « *deepfakes*²¹⁸ » ont d'abord gagné en notoriété sous la forme de vidéos d'« échanges de visage » ou dans lesquelles les mouvements des lèvres d'une personne étaient manipulés

²¹⁶ <https://openai.com/index/hello-gpt-4o/>.

²¹⁷ Johansson S., « [Scarlett Johansson's Statement About Her Interactions With Sam Altman](#) » (20 mai 2024), *The New York Times*.

²¹⁸ Le terme de « *deepfake* » est toujours utilisé dans le langage courant, mais il reste souvent associé à l'abus d'images intimes (ce qu'on appelle le « *deepfake porn* »), qui peut constituer une infraction pénale ; il s'agit d'un sujet important qui dépasse le cadre de ce chapitre.



pour les faire correspondre à d'autres paroles²¹⁹. Les *deepfakes* portant uniquement sur la voix sont également de plus en plus fréquents. Par exemple, Warner Music s'est associé aux ayants droit d'Édith Piaf pour créer un biopic de la chanteuse, qui sera raconté en utilisant sa « propre » voix grâce à la puissance de l'IA générative²²⁰. Un exemple plus controversé est celui de la désormais célèbre « *diss-track* » de Drake, *Taylor Made Freestyle*, qui comportait des voix clonées non autorisées de deux autres icônes du rap, Snoop Dogg et Tupac « 2Pac » Shakur²²¹.

Cette technologie offre des possibilités créatives remarquables. Cependant, comme l'a fait remarquer le légendaire acteur Tom Hanks, l'IA générative pose « un défi artistique, mais aussi juridique²²² », car des sosies numériques hyperréalistes peuvent être créés sans l'approbation légitime de la personne clonée. Et lorsque les œuvres, la voix ou l'image protégées d'une personne « sont utilisées à son insu, sans son consentement et sans rémunération pour générer du contenu [,] cela peut porter atteinte à ses droits moraux, économiques et de la personnalité²²³ ». C'est sur ce dernier point – les droits de la personnalité – que ce chapitre se concentrera plus particulièrement. En bref, les droits de la personnalité permettent à un individu (dans certaines circonstances et à des degrés divers) de protéger et de contrôler l'utilisation de son image ou d'autres attributs personnels (sa « personnalité »). À une époque où chacun peut reproduire l'apparence et la voix d'une autre personne rapidement, de manière convaincante et sans consentement, les droits de la personnalité font l'objet d'un débat et d'une discussion de plus en plus intenses.

De nouvelles obligations en matière de transparence sont énoncées dans le Règlement sur l'IA récemment approuvé par l'Union européenne (le « **Règlement sur l'IA**²²⁴ ») et dans la nouvelle Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit (la « **Convention**²²⁵ »). Aux États-Unis, des textes législatifs, tant au niveau des États qu'au niveau fédéral, pourraient réviser les lois sur les droits à l'image en vigueur de longue date afin de protéger les individus contre les contenus non autorisés produits par l'IA générative. Le Royaume-Uni (à la date de juillet 2024) s'est jusqu'à présent abstenu d'adopter une réglementation codifiée pour traiter spécifiquement de l'IA, et s'appuie plutôt sur son cadre de *common law* existant (précédent judiciaire), sur des dispositions législatives

²¹⁹ Lees D., (2024), « [Deepfakes in documentary film production: images of deception in the representation of the real](#) », *Studies in Documentary Film*, 18(2), p. 108–129.

²²⁰ Keslassy E., « [Creators of the Edith Piaf AI-Generated Biopic Speak Out](#) » (22 novembre 2023), *Variety*.

²²¹ Horowitz S., « [Drake retire « Taylor Made Freestyle ». Featuring AI Tupac Shakur Vocals. From Social Media After Threat of Lawsuit](#) » (28 avril 2024), *Variety*.

²²² Voir le podcast d'Adam Buxton, épisode 201 « Tom Hanks » (12 mai 2023) sur <https://shows.acast.com/adambuxton/episodes/ep201-tom-hanks>.

²²³ <https://europeanjournalists.org/blog/2023/11/23/ai-transparency-must-be-put-back-at-the-heart-of-the-ai-act/>.

²²⁴ Règlement 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 et (UE) 2019/2144 et les directives 2014/90/UE, (UE) 2016/797 et (UE) 2020/1828.

²²⁵ https://www.eeas.europa.eu/delegations/council-europe/text-first-legally-binding-global-instrument-address-risks-posed-artificial-intelligence-finalised_en.



agnostiques en matière de technologie, telles que les réglementations relatives à la protection des consommateurs, et sur le droit des contrats pour réglementer l'IA. Toutefois, on peut s'attendre à ce qu'au moins une forme de réglementation spécifique à l'IA soit introduite en temps voulu. Ce chapitre se penche sur les avancées technologiques et les pressions commerciales à l'origine de ces nouvelles lois, et se concentre sur le thème de la transparence en matière de protection des droits de la personnalité des artistes-interprètes.

5.2. Les facteurs commerciaux

5.2.1. L'évolution des doubles numériques

Malgré les progrès considérables réalisés au fil des ans dans le domaine des images générées par ordinateur et des programmes de voix synthétiques, la reproduction fidèle de l'apparence ou de la voix d'un artiste-interprète à l'aide de cette technologie héritée du passé reste quasiment impossible. Les artistes pouvaient donc jusqu'à présent raisonnablement s'attendre à conserver un minimum de contrôle et de pouvoir de négociation sur l'utilisation de leur image, impossible à imiter. Mais après l'invention des réseaux antagonistes génératifs (*generative adversarial networks*, « GAN ») en 2014, les capacités de l'IA ont rapidement été débridées et ont permis pour la première fois de générer des performances réalistes d'apparence humaine²²⁶.

Les algorithmes d'IA générative sont « entraînés » sur de vastes ensembles de données provenant de médias authentiques, tels que des vidéos d'acteurs ou des extraits d'albums de chanteurs²²⁷. Lorsqu'un utilisateur demande à l'algorithme de créer un nouveau contenu, le système d'IA utilise son apprentissage pour générer le texte, les images, le son ou la vidéo souhaités (chacun constituant un type de « sortie »). Les premières productions de l'IA générative étaient souvent peu professionnelles, avec des incohérences et des défauts visuels connus sous le nom d'« artefacts » ou d'« hallucinations », ce qui permettait d'identifier facilement les faux. Mais aujourd'hui, les résultats de l'IA générative peuvent rivaliser avec des performances authentiques grâce à des GAN de plus en plus sophistiqués et au développement ultérieur de modèles de diffusion et de transformateurs génératifs pré-entraînés (*generative pre-trained transformers* ou « GPT », comme le ChatGPT d'OpenAI). Ces progrès ont ouvert la voie à la génération réaliste, rapide et à grande échelle de contenus complexes tels que le visage et la voix humains. Les modèles de diffusion permettent notamment d'affiner les images de mauvaise qualité ou « bruyantes » jusqu'à des niveaux d'hyperréalisme, et les GPT créent des textes synthétiques qui sont à la fois nuancés et adaptés au contexte.

²²⁶ Cole S., « [AI-Assisted Fake Porn Is Here and We're ALL Fucked](#) », Motherboard Tech de VICE.

²²⁷ Ce point est particulièrement important si l'on considère les obligations de transparence énoncées à l'article 53, paragraphe 1, point d), examinées ci-dessous.



Depuis le rapport de 2020 de l'Observatoire européen de l'audiovisuel sur l'IA dans le secteur audiovisuel²²⁸, des techniques d'IA générative encore plus impressionnantes ont vu le jour. Le « *Gaussian splatting* » améliore l'éclairage, les effets d'ombre et les textures. Les « *Neural Radiance Fields* » (**NeRF**) permettent de transformer quelques *selfies* en scènes 3D complexes²²⁹. Pour les représentations artistiques, les NeRF sont en mesure de générer des expressions émotionnelles convaincantes pour les doubles numériques. L'IA générative peut également être intégrée à des logiciels plus traditionnels, notamment pour l'estimation de la pose, la photogrammétrie, la capture de mouvement et le montage vidéo.

5.2.2. Le point de vue des artistes-interprètes : autonomisation ou exploitation ?

Pendant le tournage du film de science-fiction *La Bête* (dont l'intrigue, par ailleurs, implique l'IA), l'actrice française Léa Seydoux a suggéré à l'équipe de tournage de cloner sa voix. « Je ne devrais pas travailler. Je ne devrais pas perdre de temps avec [le remplacement automatisé des dialogues²³⁰] », se souvient-elle avoir dit. Bien qu'il s'agisse d'une plaisanterie, le clonage de la voix ou de l'image d'un artiste peut certainement lui faire gagner du temps et de l'énergie, et même lui procurer de nouvelles sources de revenus. La chanteuse canadienne Grimes, connue pour sa musique électronique et son adhésion aux technologies futuristes, a annoncé qu'elle « partagerait 50 % des droits d'auteur » pour toute « chanson à succès générée par l'IA utilisant [sa] voix », et que les artistes devraient « se sentir libres d'utiliser [sa] voix sans être pénalisés²³¹ ». L'efficacité et le gain financier ne représentent qu'un volet du sujet. Le chanteur Randy Travis, membre du Country Music Hall of Fame, a perdu la capacité de parler et de chanter à la suite d'un accident vasculaire cérébral en 2012. Dix ans plus tard, il a autorisé sa maison de disques à créer une nouvelle chanson à l'aide d'une voix synthétique entraînée à partir de son catalogue²³². Présentée comme un exemple de la manière dont l'IA générative peut renforcer l'autonomie des personnes handicapées, la chanson a également d'importantes implications émotionnelles. « Depuis le jour de l'accident vasculaire cérébral, tout ce que je voulais, c'était entendre à nouveau cette voix », a fait remarquer Mary, l'épouse de Randy. « La possibilité de la retrouver est un cadeau magnifique ».

Dans ces cas, l'utilisation de l'IA générative a été approuvée par la personne dont les attributs ont été numérisés. Malheureusement, ce n'est pas toujours le cas. « Je ne vois pas d'inconvénient à ce que quelqu'un enlève un clignement d'œil lors d'un montage », a

²²⁸ Voir notamment Farish, K., « [Droits de la personnalité : d'Hollywood aux deepfakes in L'intelligence artificielle dans le secteur audiovisuel](#) » (2020), IRIS Spécial 2020-2, Observatoire européen de l'audiovisuel.

²²⁹ Voir Mildenhall, B. et al., « [NeRF : representing scenes as neural radiance fields for view synthesis](#) » (2021), *Communications of the ACM*, Volume 65, numéro 1, p. 99–106.

²³⁰ Lattanzio R., « [Léa Seydoux and George MacKay on the Darkness of L.A.](#) » (3 avril 2024), *IndieWire*.

²³¹ <https://x.com/Grimezs/status/1650304051718791170>.

²³² Carras C., « [Randy Travis releases new music with the help of AI after a stroke](#) » (7 mai 2024), *Los Angeles Times*.



expliqué un jour la star de films d'action Keanu Reeves, tout en s'insurgeant contre les *deepfakes* « effrayants » qui menacent l'autonomie des interprètes²³³. Sean Penn, vétéran d'Hollywood, est allé plus loin dans ses critiques, jugeant « insultant » et révélateur d'un « manque de moralité » le fait que les studios utilisent des doublures numériques sans le consentement de l'interprète²³⁴.

De même, Brian Cox, acteur shakespearien et vedette de *Succession*, a fustigé un studio qui, « en termes très clairs », a dit à un autre acteur qu'il conserverait les droits sur son image « et qu'il en ferait ce qu'il voudrait », ce que M. Cox a jugé « totalement inacceptable²³⁵ ». Même les partisans de l'adoption de l'IA dans l'industrie du divertissement appellent à la prudence. Reese Witherspoon, de *La Revanche d'une blonde*, a notamment admis que « l'IA devrait être un outil sur lequel nous appuyons notre propre créativité, notre propre sens de l'humanité et notre éthique. Il faut que nous ayons notre mot à dire²³⁶ ».

5.2.3. Le vide réglementaire

L'idée que les personnes devraient être à même de contrôler l'utilisation de leur nom, de leur apparence et de leur image publique constitue le fondement juridique des droits de la personnalité. Ceux-ci échappent à une définition stricte, et il peut donc être utile de les considérer comme un ensemble de droits découlant de la propriété intellectuelle, de la protection des consommateurs et de la vie privée, du droit commercial, du droit à l'image, des règles en matière de diffamation et de certains droits de l'homme²³⁷. La protection des données est un principe connexe, mais qui sert un objectif différent et, dans certains cas, les droits de la personnalité peuvent protéger des actifs immatériels qui ne sont pas des données personnelles, comme le « magnétisme de la marque » ou la réputation.

Sous réserve des circonstances factuelles et des questions de compétence, les droits de la personnalité peuvent être affirmés par plusieurs moyens, notamment par des contrats, des procédures judiciaires et des lois. En ce qui concerne les contrats, l'accord d'un musicien avec sa maison de disques pourrait fixer les limites dans lesquelles sa voix pourrait être améliorée numériquement ou clonée. Mais à l'heure actuelle, de nombreux contrats sont muets sur la question de l'IA générative, ce qui signifie qu'il peut n'y avoir en pratique aucune limite à la manière dont la maison de disques, les studios, les agences ou d'autres parties peuvent générer et distribuer des doubles numériques. En tout état de cause, certains artistes-interprètes ne disposent pas du pouvoir de négociation ou des conseils juridiques nécessaires pour protéger suffisamment leur position contractuelle.

²³³ Watercutter A., « [Keanu Will Never Surrender to the Machines](#) » (14 février 2023), *Wired*.

²³⁴ Rodrick S., « [Sean Penn's Crusade : Why He's Risking It All for Ukraine, Furious at Will Smith and Ready to Call Bulls – on Studios' AI Proposals](#) » (13 septembre 2023), *Variety*.

²³⁵ Parkel I., « [Brian Cox Rages against « Scary » AI at SAG-AFTRA Solidarity Rally in London](#) » (21 juillet 2023), *The Independent*.

²³⁶ Desborough J., « [Reese Witherspoon says artificial intelligence in Hollywood must not be feared amid actor backlash](#) » (15 avril 2024), *Mirror*.

²³⁷ Voir Farish, K., « Droits de la personnalité : d'Hollywood aux deepfakes », p. 150.



En ce qui concerne le règlement des litiges, des ressources considérables sont généralement nécessaires pour engager des procédures judiciaires ou faire des déclarations publiques visant à « nommer et dénoncer » les contrevenants pour qu'ils se conforment à la loi. Si Scarlett Johansson et les héritiers de 2Pac²³⁸ pouvaient se permettre d'engager des avocats et des experts en relations publiques, de nombreuses personnes n'ont pas accès à de tels moyens. En outre, les batailles juridiques menées dans le cadre du système judiciaire peuvent être longues, incertaines et peu efficaces du point de vue de leurs résultats, ce qui signifie qu'une partie lésée peut estimer qu'un recours en justice entraîne plus d'ennuis qu'il n'en vaut la peine.

Normalement, la législation devrait établir certains garde-fous pour les contrats et prévoir des droits minimaux, afin de réduire le risque de devoir recourir à des négociations ou à des procédures potentiellement longues et déséquilibrées. Mais les lois existantes ont largement échoué à gérer les progrès rapides et les complexités de l'IA générative, et les obligations en matière de transparence, de consentement et de responsabilité se révèlent inadéquates ou inexistantes. Ces facteurs, associés aux insuffisances potentielles du droit des contrats, ont conduit les parties prenantes concernées à exiger un nouveau cadre pour protéger efficacement les droits de la personnalité.

La chanteuse et compositrice britannique FKA Twigs l'a peut-être exprimé le mieux lors de son témoignage devant le Congrès des États-Unis en avril 2024²³⁹. « Je me servirai [de mon propre double numérique] pour accroître ma présence et gérer mes interactions avec les médias sociaux en ligne, tout en continuant à me concentrer sur mon art depuis le confort et le réconfort de mon studio », a-t-elle expliqué. Malgré son goût pour l'IA, elle a affirmé que « ce qui n'est pas acceptable, c'est que mon art et mon identité puissent simplement être récupérés par un tiers et exploités abusivement à son profit sans mon consentement, en l'absence d'un contrôle législatif approprié²⁴⁰ ». Heureusement pour FKA Twigs et les autres personnes qui partagent son opinion, un nouveau cadre législatif se profile à l'horizon.

5.3. La transparence dans les instruments européens

5.3.1. Le règlement européen sur l'IA

La transparence se caractérise par le fait de se montrer honnête, ouvert et clair au sujet d'une activité ou d'une décision particulière. Dans le contexte de l'intelligence artificielle,

²³⁸ Donahue B., « [Tupac Shakur's Estate Threatens to Sue Drake Over Diss Track Featuring AI-Generated Tupac Voice](#) » (24 avril 2024), *Billboard*.

²³⁹ « [FKA Twigs appeared before the U.S. Senate Judiciary Subcommittee on Intellectual Property on 30 April 2024 to comment on the NO FAKES Act](#) », évoqué ci-dessous.

²⁴⁰ https://www.judiciary.senate.gov/imo/media/doc/2024-04-30_-_testimony_-_twigs.pdf.



cela implique la divulgation d'informations claires et faciles d'accès sur les données utilisées pour l'entraînement, les fonctionnalités et les résultats d'un système. Grâce à ces informations, les talents, le public et les autres parties prenantes peuvent prendre des décisions éclairées sur leurs interactions avec l'IA, y compris donner ou refuser leur consentement à la création de doubles numériques. Les obligations légales de transparence peuvent donc constituer un mécanisme utile pour renforcer la légitimité et le respect des droits de la personnalité, car les individus sont alors mieux à même de comprendre et de contrôler la manière dont leur image est utilisée par ces systèmes.

Les obligations de transparence prévues par le règlement sur l'IA²⁴¹ diffèrent en fonction du profil de risque du système en question, ce qui signifie que le « type et le contenu » des règles sont adaptés à « l'intensité et à la portée des risques que les systèmes d'IA peuvent générer²⁴² ». Cette approche fondée sur le risque reflète le principe général de proportionnalité de l'Union européenne²⁴³, mais nécessite une analyse au cas par cas pour déterminer quelles obligations s'appliquent, en tenant compte de la probabilité et de la gravité des dommages potentiels²⁴⁴. Les systèmes d'IA présentant des risques « inacceptables », tels que la biométrie, le profilage ou la manipulation comportementale, sont purement et simplement interdits d'accès dans l'Union²⁴⁵. Les systèmes « à haut risque » sont ceux qui sont utilisés dans des produits ayant des implications en matière de sécurité (par exemple l'aviation ou les jouets pour enfants), ou dans des services critiques tels que l'éducation, les services financiers et juridiques, et la santé. Ces systèmes sont autorisés mais feront l'objet d'une surveillance réglementaire stricte, en raison des dommages potentiels qu'ils peuvent causer.

Les obligations de transparence pour les systèmes à haut risque sont donc importantes, mais les orientations du Parlement européen indiquent expressément que « l'IA générative, comme ChatGPT, ne sera pas classée à haut risque²⁴⁶ ». Cela dit, il serait erroné de penser que l'IA générative échappe totalement aux obligations de transparence. Au contraire, l'IA générative est considérée comme un type d'« IA à usage général » (*general-purpose AI*, « **GPAI** »), une catégorie de risque unique assortie de son propre ensemble de responsabilités en matière de transparence : à savoir, les exigences d'étiquetage du déployeur en vertu de l'article 50, et les exigences de documentation et de conformité du fournisseur en vertu de l'article 53.

Le « déployeur » est la personne physique ou morale qui utilise le système d'IA²⁴⁷. Il peut s'agir d'agences d'artistes, de sociétés de production, de maisons de disques, de créateurs individuels, etc., mais pas des personnes qui utilisent l'IA générative à des « fins

²⁴¹ Tous les « considérants » et « articles » mentionnés ci-dessous sont tirés du Règlement sur l'IA, sauf indication contraire.

²⁴² Considérant 26.

²⁴³ L'article 5, paragraphe 4, du Traité sur l'Union européenne prévoit qu'« en vertu du principe de proportionnalité, le contenu et la forme de l'action de l'Union n'excèdent pas ce qui est nécessaire pour atteindre les objectifs des traités ».

²⁴⁴ Article 3 et considérant 52.

²⁴⁵ Article 5.

²⁴⁶ <https://www.europarl.europa.eu/topics/fr/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.

²⁴⁷ Article 3.



personnelles et non professionnelles ». Le déployeur doit étiqueter les produits de l'IA générative comme étant « générés ou manipulés par une IA²⁴⁸ » et fournir cette information « d'une manière claire et reconnaissable » au plus tard lors de la « première interaction ou de la première exposition » de l'utilisateur au contenu²⁴⁹. Toutefois, et c'est là un point intéressant pour les personnes chargées des retouches et de la post-production, cette obligation ne s'applique pas lorsque l'IA est simplement utilisée dans le cadre d'« une fonction d'assistance pour la mise en forme standard²⁵⁰ », manifestement avec des logiciels comme Adobe Premiere Pro et Avid ADA.

Des exigences particulières s'appliquent aux « *deepfakes* », ou « hypertrucages » que la législation définit comme une « une image ou un contenu audio ou vidéo généré ou manipulé par l'IA, présentant une ressemblance avec des personnes, des objets, des lieux, des entités ou événements existants et pouvant être perçu à tort par une personne comme authentique ou véridique²⁵¹ ». Les utilisateurs doivent étiqueter les « *deepfakes* » comme étant générés ou manipulés artificiellement, mais là encore, une dérogation intéressante s'applique. Lorsque le *deepfake* fait partie d'un « travail ou d'un programme manifestement créatif, satirique, artistique, de fiction ou analogue », l'étiquetage peut être limité « d'une manière appropriée qui n'entrave pas l'affichage ou la jouissance de l'œuvre²⁵² ». Cette disposition offre potentiellement une grande marge de manœuvre en ce qui concerne le style et le contenu de l'étiquetage.

L'article 53 énonce les principales obligations de transparence incombant au « fournisseur » du GPAI, c'est-à-dire à la personne physique ou morale qui développe l'IA et la met ensuite sur le marché²⁵³. Ces obligations concernent principalement la tenue de registres et la documentation à mettre à la disposition des autorités de réglementation, des tiers intéressés et du public. Ce dernier point est sans doute le plus pertinent pour la protection des droits de la personnalité et exige des fournisseurs qu'ils « mettent à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner » le GPAI²⁵⁴, sur la base d'un modèle fourni par le nouveau Bureau européen de l'IA. Cette exigence vise à accroître « la transparence concernant les données utilisées dans le cadre de l'entraînement » et à « aider les parties ayant des intérêts légitimes, y compris les titulaires de droits d'auteur, à exercer et à faire respecter [leurs] droits²⁵⁵ ».

Les GPAI fournis sous « licence libre et ouverte » publient normalement leur documentation technique et les détails de leur architecture en ligne. Dans ce cas, les GPAI sous licence ouverte seront exemptés de certaines des exigences de transparence stipulées par la loi, sauf si le modèle de GPAI est considéré comme présentant un « risque important²⁵⁶ ». Cependant, dans tous les cas, les fournisseurs de GPAI sous licence ouverte

²⁴⁸ Article 50, paragraphe 2.

²⁴⁹ Article 50, paragraphe 4.

²⁵⁰ Article 50, paragraphe 2.

²⁵¹ Article 3.

²⁵² Article 50, paragraphe 4 et considérant 134.

²⁵³ Article 3.

²⁵⁴ Article 53, paragraphe 1, point d).

²⁵⁵ Considérant 107.

²⁵⁶ Considérant 104.



doivent toujours publier des résumés publics des données d'entraînement et mettre en œuvre des politiques internes pour se conformer à la loi sur le droit d'auteur.

5.3.2. La transparence dans la Convention-cadre sur l'IA

La Convention-cadre sur l'intelligence artificielle est le premier traité international juridiquement contraignant en son genre. Sous la conduite du Comité sur l'intelligence artificielle du Conseil de l'Europe, des représentants de pays tels que les États-Unis, l'Australie et le Japon ont également contribué à l'élaboration de la Convention, dont les termes peuvent être ratifiés tant par des pays européens que non européens²⁵⁷. Si un pays ratifie la Convention, sa réglementation nationale en matière d'IA doit protéger les droits de l'homme et l'État de droit, et adhérer aux principes énumérés dans la Convention²⁵⁸, notamment la transparence²⁵⁹.

La Convention stipule que la complexité, l'opacité et les différents niveaux d'autonomie des systèmes d'IA nécessitent des « garanties appropriées sous la forme de mécanismes de transparence²⁶⁰ ». Il s'agit d'« ouverture et de clarté », ce qui signifie que la logique et les détails opérationnels des algorithmes doivent être « compréhensibles et accessibles²⁶¹ ». Ce que cela signifie en pratique est toutefois sujet à interprétation, car la Convention n'impose pas le type d'obligations spécifiques et prescriptives que l'on retrouve dans le Règlement sur l'IA. Au contraire, elle est « délibérément rédigée à un niveau élevé de généralité » afin d'être « appliquée avec souplesse dans une variété de contextes en évolution rapide²⁶² ».

Pour les contenus produits par IA générative, la Convention aborde la nécessité d'éviter la tromperie et suggère des « techniques telles que l'étiquetage et le filigrane », en particulier pour les outils d'IA générative qui peuvent diffuser de la désinformation et des fausses informations²⁶³. L'accent est mis ici sur la confiance du public, la protection des consommateurs et la prévention de l'ingérence dans les processus électoraux. Malgré son importance certaine, le risque de préjudice encouru par une personne dont le double numérique apparaît dans un tel contenu n'est pas abordé, pas plus que la manière dont une personne lésée peut demander réparation.

Le principe de dignité humaine de la Convention, qui exige de « reconnaître la complexité et la richesse de l'identité [et] des émotions humaines²⁶⁴ », semble plus

²⁵⁷ Lamont C., « [The Council of Europe's draft AI Treaty : balancing national security, innovation and human rights ?](#) » (18 mars 2024) Global Governance Institute.

²⁵⁸ Tous les paragraphes mentionnés ci-dessous sont des paragraphes du [Rapport explicatif](#) de la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit.

²⁵⁹ Paragraphe 49.

²⁶⁰ Paragraphe 56.

²⁶¹ Paragraphe 57.

²⁶² Paragraphe 49.

²⁶³ Paragraphe 43.

²⁶⁴ Paragraphe 53.



rassurant. Une interprétation bienveillante de cette disposition, ainsi que l'appel de la Convention à respecter « la valeur inhérente à chaque individu », soutiennent les arguments normatifs en faveur de droits de la personnalité forts dans leur ensemble. La Convention exige une « réglementation et une gouvernance centrées sur l'humain » qui tiennent dûment compte de l'autonomie individuelle, définie dans la Convention comme la capacité d'autodétermination et de « faire des choix²⁶⁵ ». Selon cette approche, la protection de la dignité d'un individu peut théoriquement s'étendre à l'obligation d'atténuer les dommages émotionnels et psychologiques, par exemple si un double numérique est utilisé de manière diffamatoire ou non consensuelle.

En complément de la dignité humaine et de l'autonomie individuelle, la Convention énonce le principe du respect de la vie privée, qui est défini au sens large et inclut, entre autres, la protection de la « personnalité (individualité ou identité, dignité personnelle, autonomie personnelle) » et de « l'intégrité physique, psychologique ou morale ». Cette définition est tirée de l'article 8 de la Convention européenne de sauvegarde des droits de l'homme et des libertés fondamentales²⁶⁶, sur lequel la jurisprudence est abondante. Aux fins de la présente étude, la « vie privée » au sens de la dignité et de l'autonomie peut être comprise comme un droit visant à « assurer le développement, sans ingérence extérieure, de la personnalité de chaque individu dans les relations avec ses semblables²⁶⁷ ».

Pris ensemble, les principes de transparence, de dignité humaine, d'autonomie individuelle et de respect de la vie privée énoncés dans la Convention offrent des garanties contre l'utilisation abusive des doubles numériques. Bien entendu, tout dépendra en définitive de la manière dont les États membres choisiront d'interpréter et de mettre en œuvre ces dispositions dans leur législation nationale.

5.4. D'autres approches : les États-Unis et le Royaume-Uni

Berceau de Dolly Parton, Miley Cyrus, Justin Timberlake, Elvis Presley et d'innombrables autres musiciens, l'État américain du Tennessee abrite une industrie musicale dynamique, en particulier dans les régions métropolitaines de Memphis et de Nashville²⁶⁸. Il n'est donc pas surprenant que cet État ait été le premier du pays à adopter une législation spécifique visant à protéger les voix des musiciens (et les intérêts de son importante industrie du disque) contre le clonage indésirable par l'IA. En vertu de la loi ELVIS (*Ensuring Likeness Voice and Image Security*) de 2024²⁶⁹, une personne doit donner son autorisation avant que sa voix ne soit diffusée, interprétée ou rendue publique d'une autre manière²⁷⁰. La loi ELVIS introduit également une nouvelle infraction consistant à fournir « un algorithme, un

²⁶⁵ Paragraphe 55.

²⁶⁶ L'article 8 de la Convention européenne des droits de l'homme garantit que « toute personne a droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance ».

²⁶⁷ *Botta c. Italie*, Requête n° 21439/93, Cour européenne des droits de l'homme (1998).

²⁶⁸ <https://tnecd.com/wp-content/uploads/2018/10/Entertainment2015.pdf>.

²⁶⁹ <https://publications.tnsosfiles.com/acts/113/pub/pc0588.pdf>.

²⁷⁰ Section 6(a)(2).



logiciel [ou] une autre technologie » conçu pour capturer ou cloner les caractéristiques ou la voix de quelqu'un sans son consentement²⁷¹.

La Californie est un autre haut lieu de la réglementation de l'IA, avec plus de 130 propositions faites au cours de la seule session législative de 2023-2024²⁷². Parmi celles qui concernent les droits de la personnalité, on peut citer de nouvelles restrictions à l'utilisation de « répliques numériques » d'employés de l'industrie du divertissement²⁷³ et l'obligation d'apposer des « filigranes imperceptibles et hautement indélébiles » sur tout matériel créé par IA générative²⁷⁴. La loi californienne sur le droit à l'image devrait également être modifiée pour préciser qu'un double numérique est un aspect protégé des droits de la personnalité et pour obliger les systèmes d'IA générative à être assortis d'étiquettes de mise en garde à l'intention des consommateurs expliquant que leur utilisation abusive pourrait entraîner des poursuites civiles ou pénales²⁷⁵. Toutefois, les progrès risquent d'être lents, car les responsables politiques californiens sont souvent pris en étau entre les pressions exercées par les créatifs d'Hollywood, d'une part, et les innovateurs de la Silicon Valley, comme OpenAI, d'autre part.

Des travaux sont également en cours à Washington D.C. pour établir un cadre national unifié au niveau fédéral, le 118^e Congrès examinant activement plusieurs réglementations destinées à protéger les individus contre le clonage par l'IA et les « *deepfakes* » non autorisés. Il s'agit notamment de la loi « Nurture Originals, Foster Art, and Keep Entertainment Safe » (NO FAKES²⁷⁶), de la loi « No Artificial Intelligence Fake Replicas and Unauthorized Duplications » (No AI FRAUD²⁷⁷) et de la loi « Defending Each and Every Person from False Appearances by Keeping Exploitation Subject (DEEPFAKES) to Accountability²⁷⁸ ». Chaque loi relève d'une approche différente des droits de la personnalité – largement compris en Amérique comme le droit à l'image – mais toutes trois introduisent des exigences en matière de consentement et des dommages-intérêts en cas de violation.

La loi NO FAKES est une proposition bipartisanne visant à établir un droit fédéral à l'image, qui n'existe actuellement que dans certains États et avec des niveaux de protection variables²⁷⁹. Cette loi protégerait la « réplique numérique » d'un individu grâce à un nouveau droit de propriété permettant « un certain contrôle économique sur son

²⁷¹ Section 6(a)(3).

²⁷² https://leginfo.legislature.ca.gov/faces/billSearchClient.xhtml?session_year=20232024&keyword=artificial%20intelligence&house=Both&author=All&lawCode=All.

²⁷³ Projet de loi de l'Assemblée 2602, <https://legiscan.com/CA/text/AB2602/id/2928937>.

²⁷⁴ Projets de loi de l'Assemblée 3211 <https://legiscan.com/CA/text/AB3211> et 3050 <https://legiscan.com/CA/text/AB3050>.

²⁷⁵ Projet de loi du Sénat 970 <https://legiscan.com/CA/text/SB970/>.

²⁷⁶ https://www.coons.senate.gov/imo/media/doc/no_fakes_act_draft_text.pdf.

²⁷⁷ <https://www.congress.gov/bill/118th-congress/house-bill/6943/text>.

²⁷⁸ <https://www.congress.gov/bill/118th-congress/house-bill/5586/text>.

²⁷⁹ La loi du Tennessee sur la protection des droits personnels de 1984, adoptée à la suite d'un litige avec les ayants droit d'Elvis Presley, ainsi que l'article 3344 du code civil californien, sont des exemples de lois sur le droit à l'image au niveau des États.



identité²⁸⁰ », avec des exceptions pour les doubles numériques apparaissant dans des émissions d'information ou consacrées aux affaires publiques, des documentaires, des biopics, des satires, des travaux de recherche et autres. Le champ d'application de la loi No AI FRAUD est similaire, mais plus large : il accorderait à chaque individu un droit de propriété intellectuelle sur sa propre « image » et sur sa voix²⁸¹, qu'elles soient générées par l'IA ou non. Bien que la loi DEEPFAKES se concentre principalement sur les abus reposant sur des images intimes, elle introduit également plusieurs exigences en matière d'étiquetage et de divulgation afin de garantir que les médias générés par l'IA soient clairement identifiables en tant que tels.

La compatibilité de ces réglementations avec le premier amendement de la Constitution américaine, qui limite la manière dont le Gouvernement peut restreindre la liberté d'expression, est un élément de contexte clé pour leur élaboration. Dans leur version actuelle, les propositions pourraient avoir un effet dissuasif involontaire sur l'expression créative légitime, et la loi No AI FRAUD a été critiquée pour ses dispositions potentiellement « imprécises au regard de la Constitution²⁸² ».

Contrairement à l'approche européenne fondée sur les risques, le Royaume-Uni a délibérément privilégié le laisser-faire. Sa principale feuille de route en matière de législation sur l'IA a été établie dans le « livre blanc » de 2023 sur une approche pro-innovation de la réglementation de l'IA²⁸³, qui commence par affirmer qu'une législation « lourde et rigide » pourrait « étouffer l'innovation et ralentir l'adoption de l'IA ». Selon ce livre blanc, dans la mesure où le Royaume-Uni accueille « un tiers des entreprises d'IA en Europe et deux fois plus que tout autre pays européen », le Gouvernement britannique consultera les régulateurs sectoriels et les parties prenantes de l'industrie pour concevoir un cadre « proportionné » et « flexible » afin d'« alléger la charge qui pèse sur les entreprises ». Il suggère également que la législation existante, par exemple en matière de sécurité des produits et de protection des consommateurs, pourrait être suffisante pour faire face aux risques posés par l'IA.

Néanmoins, le parlement britannique semble reconnaître que l'IA générative est susceptible de créer du matériel qui « présente de manière délibérément erronée le comportement, les opinions ou le caractère d'une personne », et que certains modèles d'IA ne divulguent pas ou n'expliquent pas suffisamment leurs éléments techniques²⁸⁴. Pour lutter contre ces problèmes et d'autres, le livre blanc a introduit la notion de « transparence et d'explicabilité appropriées » parmi les cinq principes que les développeurs doivent respecter lorsqu'ils conçoivent et commercialisent des solutions d'IA. Dans sa réponse de février 2024 au livre blanc²⁸⁵, le Gouvernement a déclaré qu'il « étudiait des mécanismes permettant d'assurer une plus grande transparence, y compris

²⁸⁰ Nair P., « [Imitation Is Not Flattery : Introducing the NO FAKES Act](#) » (16 janvier 2024) ACT | The App Association.

²⁸¹ No AI FRAUD, Section 3(1) et (2).

²⁸² Klosek K., « [No Frauds, No Fakes... No Fair Use ?](#) » (1^{er} mars 2024). Association of Research Libraries.

²⁸³ <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper>.

²⁸⁴ <https://publications.parliament.uk/pa/cm5803/cmselect/cmsctech/1769/summary.html>.

²⁸⁵ <https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response>.



des mesures permettant aux titulaires de droits de mieux comprendre si le contenu qu'ils produisent est exploité par des modèles d'IA ». Toutefois, l'accent est mis ici expressément sur le droit d'auteur plutôt que sur les droits de la personnalité ou la protection de la réputation. En outre, le Gouvernement britannique reste fermement attaché à une approche non législative, de sorte que tout mécanisme de transparence de ce type serait volontaire.

Malgré tout, il est peu probable que le Royaume-Uni devienne un « Far West » dans lequel l'IA tracerait tranquillement sa route sans avoir de comptes à rendre à personne. Tout d'abord, le livre blanc admet que « l'IA exigera une action législative » en temps voulu²⁸⁶. Deuxièmement, le résultat des élections britanniques de juillet 2024 s'est traduit par l'élection d'un nouveau parti, lequel s'est prononcé sur la nécessité d'introduire un cadre réglementaire en matière d'IA²⁸⁷. Il est donc probable que l'approche décrite ci-dessus cède la place, dans ce nouveau contexte, à des obligations de transparence plus strictes et à des protections codifiées pour les individus.

5.5. La transparence, pilier du respect des droits de la personnalité

Bien entendu, la législation ne représente qu'une partie de la saga des droits de la personnalité à l'ère de l'IA. De nombreuses entreprises pratiquent l'autorégulation, avec des groupes tels que la Coalition for Content Provenance and Authenticity (C2PA²⁸⁸) et le National Institute of Standards and Technology (NIST²⁸⁹), qui établissent des normes techniques et de gouvernance volontaires. La pression publique et les groupements d'intérêt ont également un impact, comme en témoigne la résolution relative au mouvement de grève de la SAG-AFTRA en décembre 2023²⁹⁰. Après cinq mois de conflit social, le syndicat des travailleurs du monde du divertissement a approuvé un accord avec l'Alliance of Motion Picture and Television Producers (AMPTP²⁹¹), la principale association professionnelle américaine représentant les studios de cinéma, les chaînes de télévision, les services de diffusion en continu et les sociétés de production. Entre autres choses, les membres de l'AMPTP doivent désormais obtenir le consentement de l'artiste-interprète lorsqu'ils réalisent des doubles numériques. En juin 2024, il y avait également plus de 20 procès en cours en matière de propriété intellectuelle de part et d'autre de l'Atlantique impliquant des entreprises d'IA générative, dont l'issue influencera vraisemblablement la manière dont elle sera développée et utilisée²⁹².

²⁸⁶ *Ibid.*

²⁸⁷ Voir, entre autres, Landi, M. « *Labour commits to introducing AI regulation for tech giants* », *The Independent* (13 juin 2024) et le Manifeste du parti travailliste.

²⁸⁸ <https://c2pa.org/specifications/specifications/2.0/index.html>.

²⁸⁹ <https://www.nist.gov/itl/ai-risk-management-framework>.

²⁹⁰ <https://www.sagaftra.org/sag-aftra-members-approve-2023-tvtheatrical-contracts-tentative-agreement>.

²⁹¹ [Résumé de l'accord mis à jour](#).

²⁹² Lee E., « *Status of all 24 copyright lawsuits v. AI companies* » (24 mai 2024), *ChatGPTiseatingtheworld*.



La question de savoir comment la législation, les initiatives émanant de l'industrie, les négociations contractuelles et la jurisprudence évolueront pour relever les défis posés par l'IA générative n'est pas encore tranchée. Ce qui est clair, cependant, c'est que la transparence est cruciale pour un exercice véritable des droits de la personnalité, alors que les doubles numériques et le clonage par l'IA deviennent de plus en plus courants. Le consentement est crucial au contrôle, et sa légitimité passe par l'accès à des informations honnêtes sur les risques et les avantages de l'IA générative. De plus, le fait que les fournisseurs et les déployeurs fassent preuve de transparence sur la manière dont l'IA générative est développée et utilisée contribue à faire en sorte qu'ils n'échappent pas à leurs responsabilités et que les personnes lésées disposent de moyens de recours adéquats. Cela peut protéger les artistes-interprètes, les créateurs et le public, et favoriser la confiance dans une utilisation légitime des doubles numériques et des systèmes qui les créent, dans le respect de la loi. Plus important encore peut-être, le fait que les personnes soient pleinement informées de la manière dont leur image, leur voix ou d'autres attributs personnels sont numérisés contribue à affirmer leurs droits à l'autodétermination, à la dignité et à l'autonomie. Pour reprendre les mots de l'actrice Talulah Riley : « Il est vital que ma voix et mon image m'appartiennent, peu importe qu'il soit facile et peu coûteux de les reproduire numériquement²⁹³ ».

²⁹³ Vallance C., « [Actors launch campaign against AI “show stealers”](#) » (21 avril 2022), *BBC*.



6. Impact de l'IA sur le marché du travail audiovisuel en Europe

Elodie Migliore, Doctorante au sein du CEIPI, Université de Strasbourg

6.1. Introduction

« Toute puissance est faible, à moins que d'être unie ».

Le vieillard et ses enfants – Jean de La Fontaine.

L'essor des technologies de l'IA a un impact sur tous les aspects de notre vie quotidienne. Comme l'illustrent de nombreuses œuvres de science-fiction, telles que *I, Robot*²⁹⁴, la perception commune des technologies de l'IA est qu'elles remplaceront les humains dans de nombreuses tâches. Bien que certains y voient une chimère, cette prédiction devient déjà réalité dans certains secteurs, où ces technologies se substituent effectivement au travail humain.

Le milieu de la création en est un exemple frappant. Une étude préliminaire d'OpenAI indique que le risque d'exposition à l'IA pour les écrivains et les auteurs est de 82,5 %²⁹⁵. Le secteur audiovisuel ne fait pas exception. Un rapport de KPMG²⁹⁶ souligne que les professions créatives sont celles qui comptent le plus grand nombre de tâches susceptibles d'être automatisées, avec une part de 43 % de tâches automatisées pour les auteurs, les rédacteurs et les traducteurs, les humains se contentant de « peaufiner » la production des machines²⁹⁷. Cela pourrait avoir de nombreuses conséquences, comme la réduction des salaires des auteurs professionnels. Cependant, les travailleurs ont décidé de ne pas rester les bras croisés. C'est ainsi que des grèves et des mouvements importants ont eu lieu au cours des derniers mois.

²⁹⁴ *I, Robot*, Alex Proyas, 20th Century Fox, 2004.

²⁹⁵ Tyna Eloundou et autres, « [GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models](#) » (*arXiv*, 21 août 2023).

²⁹⁶ [Generative AI and the UK Labour Market](#), KPMG Royaume-Uni.

²⁹⁷ [Writers and AI](#) (Writers' Guild of Great Britain, 12 juillet 2023).



Ce chapitre vise à analyser l'état actuel du droit du travail dans le secteur audiovisuel en ce qui concerne l'utilisation de l'IA, en revenant sur les deux grandes grèves qui ont eu lieu aux États-Unis (US).

6.2. Impact de l'IA sur le droit du travail dans le secteur audiovisuel aux États-Unis

6.2.1. Les grèves WGA et SAG-AFTRA

Le 2 mai 2023, la Writers Guild of America (WGA), un syndicat représentant 11 500 scénaristes, s'est mise en grève. L'ordre de grève concernait toutes les sociétés signataires de l'accord de base minimum (*Minimum Basic Agreement*, MBA²⁹⁸), une convention collective qui fixe les règles et les taux de rémunération applicables aux scénaristes de la WGA²⁹⁹. Les scénaristes n'avaient pas fait grève depuis le mouvement historique de 100 jours en 2007³⁰⁰.

Dans la foulée, le 14 juillet 2023, le Screen Actors Guild-American Federation of Television and Radio Artists (SAG-AFTRA), un syndicat représentant 160 000 professionnels des médias dans le monde entier, s'est également mis en grève. Le mot d'ordre de grève concernait tous les services couverts par le *Producer SAG-AFTRA Codified Basic Agreement*, les *SAG-AFTRA Television Agreements* et les accords connexes³⁰¹. C'était la première fois que les acteurs participaient à un conflit social aux États-Unis depuis la grève des acteurs de 1980³⁰². En outre, pour la première fois depuis 1960, les acteurs et les scénaristes étaient simultanément en grève.

Les deux syndicats s'opposaient à l'Alliance of Motion Picture and Television Producers (AMPTP), une association professionnelle représentant 350 sociétés de production télévisuelle et cinématographique américaines dans le cadre de négociations collectives avec les syndicats de l'industrie du divertissement.

La grève de la WGA a pris fin après 146 jours, le 27 septembre 2023, à la suite d'un accord conclu avec l'AMPTP, couvrant la période du 25 septembre 2023 au 1^{er} mai 2026.

²⁹⁸ [Memorandum of Agreement for the 2023 WGA Theatrical and Television Basic Agreement](#), 2023.

²⁹⁹ Elodie Migliore, « [Fin de la grève des scénaristes américains : quand l'union fait la force](#) », *Intelligence artificielle, Dalloz Actualité* (2023).

³⁰⁰ Cal Berry, « [Blueprint for a Strike in the Entertainment Industry: Lessons from the 2007 WGA Strike](#) », (*Left Voice*, 5 novembre 2021). Voir aussi, « [Pencils Down! The 100 Days of the Writers Guild Strike](#) », Brian S. Kalata, 2014.

³⁰¹ [SAG-AFTRA Strike Order for TV/Theatrical/Streaming Contracts](#) (SAG-AFTRA).

³⁰² Cynthia Littleton, « [Revisiting the 1980 SAG-AFTRA Strike with "MASH" Stars, an Emmy Boycott and All-Night Negotiating Sessions: "We'Re Going to Strike Like Hell"](#) », (*Variety*, 1^{er} septembre 2023).



La grève SAG-AFTRA s'est terminée le 9 novembre 2023 avec un accord ratifié le 5 décembre, couvrant la période du 9 novembre 2023 au 30 juin 2026³⁰³.

Les deux grèves étaient motivées par des revendications communes, telles que la négociation de nouveaux droits résiduels liés à l'essor des services de diffusion en continu, mais elles avaient également pour objectif commun la réglementation de l'utilisation de l'IA générative dans le cadre professionnel.

Les scénaristes de la WGA craignaient la concurrence des œuvres générées par l'IA sur leurs emplois et la diminution de leur statut et de leurs droits résiduels du fait de l'entraînement des modèles d'IA à partir de matériel produit par des scénaristes professionnels. La SAG-AFTRA redoutait que les studios n'utilisent l'IA et les technologies numériques pour reproduire les visages et les voix des artistes-interprètes, réduisant les droits et les possibilités de travail des acteurs.

6.2.2. L'accord WGA après la grève

Les revendications de la WGA étaient nombreuses. Si certaines d'entre elles ont été acceptées, d'autres n'ont pas été intégrées dans l'accord en l'état³⁰⁴.

Tout d'abord, l'utilisation de l'IA générative n'est pas autorisée pour écrire ou réécrire du matériel ou du contenu littéraire. En outre, le contenu généré par l'IA ne peut pas être considéré comme du matériel source dans le cadre de l'accord³⁰⁵.

En outre, un scénariste peut utiliser un système d'IA dans le cadre de ses services si l'entreprise y consent, à condition que le scénariste respecte les conditions fixées par celle-ci. Toutefois, une société ne peut pas imposer à un scénariste l'utilisation d'un système d'IA pour fournir ses services. L'entreprise peut également refuser l'utilisation d'un système d'IA, notamment si elle a des doutes sur la possibilité de bénéficier de la protection du droit d'auteur pour le contenu produit, ou sur sa capacité à exploiter ce contenu. L'entreprise doit également informer l'auteur si les documents qui lui sont communiqués ont été générés par un système d'IA ou contiennent des éléments générés par un système d'IA.

Enfin, la question de l'entraînement des systèmes d'intelligence artificielle, auquel la WGA était fermement opposée, a fait l'objet d'une controverse. Cette question a été l'une des plus difficiles à régler, et l'accord a maintenu une clause selon laquelle, si les auteurs conservaient des droits réservés sur leur matériel, ils pouvaient – ou la WGA en leur nom – interdire l'utilisation de ce matériel pour l'entraînement d'un système d'IA

³⁰³ [2023 TV/Theatrical Contracts Tentative Agreement](#).

³⁰⁴ Article 72, page 68, WGA, Proposition N° 29, protocole d'accord sur l'accord de base pour le cinéma et la télévision de 2023 (*Memorandum of Agreement for the 2023 WGA Theatrical and Television Basic Agreement*).

³⁰⁵ « On entend par matériel source tout le matériel sur lequel le scénario est basé, autre que l'histoire telle que définie ci-dessus, notamment tout autre matériel sur lequel l'histoire est basée. Le générique doit indiquer à l'écran l'auteur de l'histoire d'un long métrage [...] », voir [l'accord de base minimum de la WGA de 2020](#), p. 403. Le contenu généré par l'IA ne peut être utilisé pour porter atteinte au travail ou aux droits d'un scénariste.



généraliste. Inversement, cela signifie également que si un studio conserve pleinement les droits réservés sur le matériel, il peut l'exploiter pour entraîner des systèmes d'IA générative³⁰⁶. Cette clause n'est donc pas aussi avantageuse qu'il n'y paraît, car il n'est pas interdit aux studios d'utiliser des scénarios dont ils sont propriétaires pour entraîner des systèmes d'IA ; tout dépendra des droits conservés.

6.2.3. L'accord SAG-AFTRA

Si l'accord WGA constitue une avancée en matière de réglementation de l'IA dans le secteur audiovisuel, l'accord SAG-AFTRA intègre également des dispositions intéressantes³⁰⁷. Le protocole d'accord SAG-AFTRA de 2023 s'applique aux films cinématographiques et aux contenus dramatiques scénarisés produits pour la télévision et les nouvelles plateformes médiatiques, et se concentre plus particulièrement sur les questions relatives à la production³⁰⁸. Selon l'accord, le consentement et la rémunération sont deux notions fondamentales, qui reviennent dans toutes les dispositions relatives à l'IA.

6.2.3.1. Données d'apprentissage

Tout d'abord, en ce qui concerne la question des données d'apprentissage, il apparaît que le protocole d'accord de la SAG-AFTRA ne prévoit pas de compensation supplémentaire pour l'inclusion de séquences ou d'enregistrements vocaux de la performance d'un acteur dans un ensemble de données d'apprentissage. Cela ne signifie pas qu'il serait impossible de demander une compensation, mais qu'il appartient aux acteurs ou aux interprètes de négocier leurs propres accords. Cela semble possible pour les acteurs célèbres disposant d'un pouvoir de négociation suffisant, mais moins réaliste pour les nouveaux acteurs entrant sur le marché.

La seule disposition traitant des données d'apprentissage est le paragraphe C du titre II "Intelligence artificielle" du résumé de l'accord provisoire de 2023 succédant à l'accord de base codifié de 2020 entre les producteurs et la SAG-AFTRA et à l'accord de télévision de 2020 de la SAG-AFTRA³⁰⁹, qui prévoit des réunions régulières pour « [...] discuter de la rémunération, le cas échéant, pour l'utilisation du travail produit dans le cadre de [la convention collective] afin de former le système d'IA générative à la création d'interprètes synthétiques [...]»³¹⁰.

³⁰⁶ Article 72, page 68, WGA Proposition n°29, protocole d'accord sur l'accord de base pour le cinéma et la télévision de 2023.

³⁰⁷ Toutefois, il existe des opinions dissidentes, voir à ce sujet Laura Weiss, « [SAG-AFTRA's New Contract Falls Short on Protections from AI](#) » (*Prism*, 5 décembre 2023). Voir aussi « [How SAG-AFTRA's AI Provisions Work : A Lawyer's View](#) ».

³⁰⁸ *2023 TV/Theatrical Contracts Tentative Agreement*, *op. cit.*

³⁰⁹ https://www.sagaftra.org/files/sa_documents/TV-Theatrical_23_Summary_Agreement_Final.pdf

³¹⁰ *2023 TV/Theatrical Contracts Tentative Agreement*, Section C, page 3, *op. cit.*



6.2.3.2. Interprètes synthétiques

Deuxièmement, l'accord définit le concept d'interprète synthétique comme un élément généré numériquement qui « est destiné à créer, et crée effectivement, la nette impression que l'élément est un interprète naturel qui n'est pas reconnaissable comme un interprète naturel identifiable ; dont la voix n'est pas celle d'une personne physique ; qui n'est pas non plus un double numérique ; et pour lequel il n'existe aucun contrat de travail pour le film avec un interprète naturel dans le rôle interprété par l'élément³¹¹ ».

Il prévoit ensuite des exigences supplémentaires pour l'utilisation d'interprètes synthétiques reconnaissables. Cette notion fait référence aux artistes-interprètes synthétiques comprenant des caractéristiques reconnaissables d'un acteur, telles qu'une « caractéristique faciale principale (c'est-à-dire les yeux, le nez, les oreilles et/ou la bouche) » qui est créée par le biais d'un « *prompt* à un système d'IA générative³¹² ». Dans ce cas, le producteur doit négocier et obtenir le consentement de l'artiste-interprète. Par exemple, si quelqu'un souhaite générer un artiste-interprète synthétique avec les yeux d'Emma Stone, il lui faudra négocier et obtenir son consentement explicite.

6.2.3.3. Doubles numériques

Troisièmement, l'accord définit deux types de doubles, les doubles basés sur l'emploi et les doubles indépendants³¹³.

Les doubles numériques basés sur l'emploi sont définis comme des reproductions numériques de la voix ou de l'image d'un artiste-interprète créées dans le cadre de son travail sur un film, à l'aide de technologies numériques et de la participation physique de l'artiste-interprète, pour le représenter dans des prises de vue ou des bandes sonores où il n'a pas réellement joué³¹⁴. À titre d'exemple, on peut citer la création d'une réplique de Kyle MacLachlan pour représenter le jeune Henry MacLean dans *Fallout*³¹⁵.

Dans ce cas, le producteur doit avertir l'acteur avant la création du modèle et obtenir son consentement « clair et manifeste³¹⁶ » dans un document distinct du contrat de travail signé par l'artiste-interprète, assorti de dispositions supplémentaires en matière de rémunération. L'accord comporte ensuite une section entière consacrée à l'utilisation de ces doubles, précisant les conditions dans lesquelles un consentement ou une

³¹¹ 2023 TV/Theatrical Contracts Tentative Agreement, Section C, page 4, *op. cit.*

³¹² *Ibid.*

³¹³ *Ibid.*

³¹⁴ *Ibid.*

³¹⁵ Pour un double numérique qui n'est pas du goût de tout le monde et qui a fait l'objet de discussions entre les fans, certains d'entre eux avançant l'hypothèse qu'il avait été généré par l'intelligence artificielle et mentionnant les accords SAG-AFTRA, voir par exemple cette [discussion](#) sur *Reddit*.

³¹⁶ 2023 TV/Theatrical Contracts Tentative Agreement, Section C, *op. cit.*



rémunération supplémentaire est nécessaire³¹⁷. Une « description suffisamment précise de l'utilisation prévue³¹⁸ » doit également être incluse.

Les doubles numériques créées de manière indépendante sont des répliques conçues pour représenter de manière convaincante un artiste-interprète naturel en utilisant des caractéristiques reconnaissables telles que sa voix et/ou son image. Le double sera utilisé pour interpréter un personnage plutôt que l'artiste-interprète naturel et il n'y a pas de contrat de travail avec l'artiste-interprète naturel pour le film dans lequel le double est utilisé. La réplique est souvent créée en utilisant du matériel existant pour représenter l'acteur dans des scènes qu'il n'a pas réellement tournées³¹⁹, par exemple Paul Walker dans le rôle de Brian O'Conner dans *Fast and Furious 7*. Pour ce type de réplique, le producteur doit négocier et obtenir le consentement de l'acteur avant de l'utiliser. Il couvre également les cotisations de retraite et d'assurance-maladie.

6.2.3.4. La modification numérique

Enfin, l'accord traite également du concept de modification numérique, un phénomène courant dans l'industrie cinématographique. La modification numérique peut être effectuée à des fins cosmétiques, par exemple, et n'implique pas toujours des processus d'IA. Le consentement ne sera pas nécessaire lorsque « les images ou les enregistrements sonores de l'artiste interprète restent pour l'essentiel conformes au scénario, à l'interprétation et/ou à l'enregistrement », mais le consentement sera nécessaire pour des modifications plus importantes. Des règles similaires concernant le consentement et la rémunération sont incluses.

6.3. Impacts de l'IA sur le droit du travail dans le secteur audiovisuel dans l'UE

6.3.1. L'approche de l'Union européenne

Les deux grèves aux États-Unis ont permis d'améliorer les conditions de travail dans le secteur audiovisuel. L'Union européenne (UE) commence également à se pencher sur ces questions.

Tout d'abord, il convient de noter que la politique sociale relève principalement de la responsabilité des États membres de l'UE, ce qui pourrait avoir un impact sur la capacité de l'UE à renforcer les droits des travailleurs dans le cadre du recours à l'IA.

³¹⁷ L'accord contient également des dispositions concernant les acteurs décédés ou l'utilisation de ces doubles pour une suite ou un préquel par exemple, voir *ibid*, page 5.

³¹⁸ *2023 TV/Theatrical Contracts Tentative Agreement*, Section C, *op. cit.*

³¹⁹ *Ibid.*



Toutefois, certains domaines relèvent d'une compétence partagée avec l'UE³²⁰. En effet, le titre X du Traité sur le fonctionnement de l'Union européenne (TFUE³²¹) définit la politique sociale dans l'UE³²². Une clause sociale horizontale est également introduite par l'article 9 du TFUE. Enfin, l'article 6 du traité sur l'Union européenne (TUE)³²³ attribue un pouvoir contraignant aux droits sociaux énoncés dans la Charte des droits fondamentaux de l'Union européenne.

Le Parlement européen et le Conseil peuvent adopter des mesures d'incitation pour soutenir et compléter les actions des pays de l'UE dans des domaines spécifiques. Ils peuvent également adopter des exigences minimales par le biais de directives afin de permettre aux pays de l'UE d'adopter des dispositions supplémentaires plus strictes. Ces directives ne concernent que certains domaines, notamment la santé et la sécurité des travailleurs, l'information et la protection des travailleurs ou la protection des travailleurs en cas de résiliation de leur contrat de travail³²⁴.

Cela signifie que la Commission européenne a une compétence limitée en matière sociale, notamment en ce qui concerne la rémunération, ce qui explique pourquoi elle pourrait ne pas agir aussi vite dans ces domaines.

Pour l'instant, il n'existe pas de législation européenne contraignante spécifiquement axée sur le secteur audiovisuel et imposant de nouvelles conditions aux États membres. Néanmoins, une avancée peut être observée au Parlement européen avec sa résolution du 21 novembre 2023, qui formule des recommandations à la Commission sur un cadre de l'UE relatif à la situation sociale et professionnelle des artistes et des travailleurs des secteurs de la culture et de la création³²⁵.

La résolution est composée de 73 recommandations, comprenant une grande variété de mesures relatives au statut de l'artiste, à la protection sociale, à la décence des conditions de travail, à la rémunération équitable, à l'éducation et à la formation, à la liberté artistique ou à la négociation collective.

La résolution appelle à une approche stratégique au niveau de l'Union pour aborder l'impact sur les emplois, les méthodes de travail, les conditions de travail et la nécessité de développer les compétences et les possibilités de reconversion professionnelle. Elle recommande également des plans de création d'emplois et un soutien financier sectoriel pour garantir la protection sociale des personnes touchées par

³²⁰ La compétence partagée concerne les domaines dans lesquels la législation et l'adoption d'actes contraignants peuvent relever à la fois du niveau européen et de chacun des États membres, indépendamment des autres. Toutefois, les États membres ne peuvent exercer leur compétence que dans la mesure où l'UE n'a pas exercé, ou a décidé de ne pas exercer la sienne. En ce qui concerne la politique sociale, elle ne concerne que les aspects spécifiquement définis dans le traité.

³²¹ [Versions consolidées du Traité sur l'Union européenne et du Traité sur le fonctionnement de l'Union européenne 2012/C 326/01.](#)

³²² Les objectifs de la politique sociale sont précisés à l'article 151 du TFUE.

³²³ [Version consolidée sur Traité sur l'Union européenne C 202/1.](#)

³²⁴ « Politique sociale » Glossaire des synthèses, *EUR-Lex*.

³²⁵ [Résolution du Parlement européen du 21 novembre 2023 contenant des recommandations à la Commission sur un cadre de l'Union pour la situation sociale et professionnelle des artistes et des travailleurs des secteurs de la culture et de la création \(2023/2051\(INL\)\).](#)



la numérisation et les pertes d'emplois liées à l'IA. La résolution met l'accent sur des défis liés à la perte d'emploi et à la transformation du travail et exhorte la Commission à tenir compte des effets de l'IA sur les droits et le bien-être des travailleurs dans la législation future.

La Commission européenne s'est félicitée de la résolution³²⁶ et a souligné certaines initiatives déjà prises dans ce domaine³²⁷. Un groupe de haut niveau est prévu pour étudier la meilleure façon de répondre aux besoins du secteur. La Commission a conclu sa réponse en énumérant plusieurs initiatives, telles que la rémunération équitable des auteurs et des artistes-interprètes pour l'exploitation de leur travail artistique, en mettant l'accent sur les défis posés par l'IA.

6.4. Analyse d'un éventail d'initiatives émanant de différents acteurs

Les sections suivantes examinent le rôle joué par différents acteurs essentiels du secteur audiovisuel, tels que les organismes de gestion collective, les associations ou fédérations et les syndicats, et étudient leurs actions et initiatives.

6.4.1. Les organismes de gestion collective (OGC)

Parmi les premières entités qui viennent à l'esprit lorsqu'il s'agit d'améliorer les conditions de travail dans le secteur audiovisuel, se trouvent les organismes de gestion collective (OGC)^{328,329}. Par conséquent, la SACD et Playright seront analysées en raison de la nature de leurs actions, ainsi que la SAA, une association regroupant différents OGC³³⁰.

Tout d'abord, la Société des Auteurs et Compositeurs Dramatiques française (SACD) a adapté ses contrats généraux avec les utilisateurs en y incluant de nouvelles clauses visant à protéger les œuvres de ses auteurs. Ces clauses empêchent les utilisateurs de concéder les droits des auteurs qu'elle représente à des sociétés d'IA³³¹. Par

³²⁶ [Réponse](#) de la Commission européenne à la résolution du 21 novembre 2023.

³²⁷ Commission européenne, Direction générale de l'éducation, de la jeunesse, du sport et de la culture, [Le statut et les conditions de travail des artistes et des professionnels de la culture et de la création – Rapport du groupe de travail « Méthode ouverte de coordination » \(MOC\) composé d'experts des États membres](#), Office des publications de l'Union européenne, 2023.

³²⁸ Les organismes de gestion collective visent à fournir aux auteurs un moyen efficace et rentable de gérer leurs droits dans le monde entier afin de garantir que leurs œuvres sont utilisées conformément aux lois en vigueur.

³³⁰ Même si cette partie se concentre sur trois acteurs, il existe une grande variété d'OGC, telles que le BECS, le GVL ou l'AIE, qui agissent également dans le domaine de l'IA.

³³¹ SAA, [Séminaire d'experts sur les droits des auteurs audiovisuels et l'intelligence artificielle](#), 30 janvier 2024. Voir également [5 enseignements tirés du séminaire d'experts de la SAA sur l'intelligence artificielle](#), 30 janvier 2024.



exemple, le modèle de contrat régissant la relation entre producteur et auteur contient plusieurs dispositions similaires au protocole d'accord de la WGA³³².

Il contient des clauses empêchant de forcer les auteurs à utiliser des systèmes d'IA, une obligation de divulgation de l'utilisation d'un système d'IA par l'auteur ou par le producteur à l'autre partie, ou encore des clauses imposant au producteur de ne pas utiliser d'éléments générés par l'IA pour créer des visuels destinés à la promotion du film sans l'accord explicite et préalable de l'auteur. De plus, contrairement aux accords WGA et SAG-AFTRA, l'utilisation des œuvres produites par l'auteur pour entraîner un système d'IA est expressément interdite. Cependant, la portée de ces contrats types sera limitée par rapport à ce qui a été mis en place aux États-Unis, car leur utilisation n'est que conditionnelle. Ils ne constituent pas une base minimale obligatoire à utiliser et, dans la pratique, c'est à l'auteur, par son pouvoir de négociation, qu'il revient d'imposer ces conditions contractuelles à l'autre partie. Enfin, une nouvelle clause a été soumise à l'assemblée générale de la SACT, qui permettrait à la SACT « d'intervenir sans équivoque pour le compte de ses auteurs membres sur l'intelligence artificielle³³³ ». Le rapport annuel 2023 évoque également plusieurs questions qui doivent être posées, telles que la manière de parvenir à un accord avec les sociétés d'intelligence artificielle, comme elles l'ont fait avec tous les diffuseurs et plateformes³³⁴.

PlayRight, un OGC belge, a fait part de son opinion sur l'IA dans un document de synthèse ou un communiqué sur l'IA³³⁵. Elle établit également des lignes directrices pour la conclusion de contrats avec l'IA³³⁶, sans fournir de modèle de contrat, mais des recommandations sur les éventuelles clauses à inclure dans les accords contractuels. Certaines recommandations sont également similaires aux dispositions des accords WGA et SAG-AFTRA³³⁷.

Enfin, les initiatives de la Société des auteurs de l'audiovisuel (SAA) méritent d'être mentionnées. La SAA est une association d'organismes européens de gestion collective représentant les auteurs audiovisuels. La SAA plaide activement en faveur d'une réglementation de l'utilisation de l'IA par le biais de déclarations communes ou de prises de position³³⁸. Elle sensibilise également aux questions liées à l'IA au moyen de séminaires, par exemple³³⁹. Elle fournit donc un large éventail d'informations sur l'IA et son impact sur le secteur audiovisuel³⁴⁰. En outre, elle a récemment publié un document

³³² [Modèles de contrats audiovisuels](#), SACT, 8 février 2017.

³³³ Voir [Rapport annuel de 2023](#), SACT, 2024.

³³⁴ *Ibid*, page 10.

³³⁵ Playright, [Mémoire politique de l'artiste-interprète 2024-2029](#), 16 mai 2024.

³³⁶ Team COMPLUS, [Guidelines contractuelles relatives à l'IA](#), PlayRight, 27 février 2024.

³³⁷ Par exemple, ils examinent la possibilité d'inclure une clause contractuelle spécifique exigeant l'autorisation préalable de l'artiste-interprète pour toute nouvelle utilisation, Playright, [Guidelines contractuelles relatives à l'IA](#), *op. cit.* Voir également Team COMPLUS, [Fin de la grève à Hollywood : SAG-AFTRA conclue un accord](#), PlayRight, 28 novembre 2023.

³³⁸ Voir [Règlement de l'UE sur l'IA : déclaration commune des créateurs et détenteurs de droits européens](#), SAA, 13 mars 2024.

³³⁹ Voir par exemple, [Séminaire d'experts sur les droits des auteurs audiovisuels et l'intelligence artificielle](#), *op. cit.*

³⁴⁰ Voir leurs différentes actions concernant l'IA <https://www.saa-authors.eu/en/tags/222-ai#.Zlm3Y3bP06Q>.



de position³⁴¹ axé sur les questions de propriété intellectuelle et soulignant les principes fondamentaux d'une réglementation de l'IA centrée sur l'humain qui encourage la créativité³⁴². Ce document établit des principes de transparence, d'autorisation/licence et de rémunération, tout en soulignant le rôle que peuvent jouer les OGC³⁴³.

6.4.2. Associations et fédérations

Les associations et fédérations européennes ou internationales tendent également à converger vers des actions communes pour avancer. Tout d'abord, il est important de noter que les associations européennes n'ont peut-être pas le même pouvoir de négociation que leurs homologues américaines en raison de leur fragmentation géographique et de leurs différences structurelles. Cependant, elles œuvrent activement à l'amélioration des conditions de travail des travailleurs³⁴⁴.

Un exemple notable est la Fédération des scénaristes d'Europe (FSE), qui regroupe 26 organisations de scénaristes. Les actions de cette fédération sont intéressantes car elle représente les travailleurs à risque. Son objectif est notamment d'améliorer les conditions de travail des travailleurs par le biais de campagnes de sensibilisation et d'activités de représentation³⁴⁵. Elle a publié une résolution commune avec l'International Affiliation of Writers Guilds (IAWG) appelant à une « utilisation éthique » de l'IA sur la base de lignes directrices³⁴⁶. En outre, l'une des principales priorités de la FSE est de faciliter les négociations collectives, tant au niveau de l'UE qu'aux niveaux national et régional. La négociation collective permet d'établir des conditions minimales d'emploi. C'est l'une des forces des syndicats observés aux États-Unis, qui leur a permis de parvenir à de tels accords. Une approche similaire pourrait s'avérer utile pour conclure de nouveaux accords encadrant l'IA³⁴⁷.

La FSE est loin d'être la seule association ou fédération à agir pour améliorer les conditions des travailleurs au regard de l'utilisation de l'IA dans le secteur audiovisuel. Par exemple, la Fédération européenne des réalisateurs de l'audiovisuel (FERA), l'Association européenne de production audiovisuelle (CEPI) ou l'Association des organisations européennes d'artistes-interprètes (AEPO-ARTIS) participent à cette action

³⁴¹ [Artificial intelligence must serve society and enhance human creativity](#), SAA, 4 octobre 2023.

³⁴² *Ibid*, page 5.

³⁴³ *Ibid*.

³⁴⁴ Voir par exemple, Fédération européenne des réalisateurs de l'audiovisuel (FERA), Association européenne de production audiovisuelle (CEPI), Association des organisations européennes d'artistes-interprètes (AEPO-ARTIS) ou la SAA.

³⁴⁵ Voir par exemple les nombreuses lettres conjointes appelant à réglementer l'IA ou concernant les réglementations relatives à l'IA : [Joint letter in response to the dialogue with the Audiovisual Sector on Copyright & AI](#), FSE, 30 novembre 2023, [For an innovation and creator friendly AI Act](#), FSE, 24 novembre 2023 et [Joint Statement on Artificial Intelligence and the Draft EU AI Act](#), FSE, 26 septembre 2023.

³⁴⁶ Fédération des scénaristes d'Europe, [Artificial Intelligence : Global Screenwriters Call for Ethical Use](#), FSE, 11 avril 2024. Voir aussi, [Global Screenwriters Call for AI Regulation](#).

³⁴⁷ Fédération des scénaristes d'Europe, [Collective Bargaining](#), FSE.



par le biais de lettres conjointes, d'événements ou en rendant publiques des informations sur le sujet.

La Confédération internationale des sociétés d'auteurs et compositeurs (CISAC), l'une des organisations à but non lucratif les plus importantes du secteur, se concentre également sur le lobbying par le biais de lettres conjointes³⁴⁸, de mesures de communication³⁴⁹, d'éducation³⁵⁰ et de diffusion d'informations. Par exemple, dans son récent rapport annuel³⁵¹, elle a souligné que l'IA était l'une de ses principales priorités, en se référant à trois principes, qui ne portent pas nécessairement sur des questions de droit du travail, mais plutôt sur l'autorisation d'utiliser les œuvres des artistes³⁵², leur rémunération et la transparence³⁵³.

6.4.3. Organisations syndicales

Les syndicats semblent jouer un rôle important pour définir l'encadrement de l'IA, comme cela a été constaté aux États-Unis.

UNI Europa est une fédération syndicale européenne qui tire parti de sa puissance pour développer la négociation collective. UNI MEI est sa section audiovisuelle³⁵⁴. UNI Europa a adopté une résolution sur l'IA en 2021, plaidant pour des salaires et des conditions de travail équitables par le biais de la négociation collective³⁵⁵. Cette résolution établit une « plateforme politique tournée vers l'avenir pour un cadre commun sur les revendications en matière de négociation collective³⁵⁶ ». Cependant, il faut garder à l'esprit que les salaires et les conditions de travail sont principalement déterminés par les différents cadres juridiques et institutionnels nationaux. La négociation collective relève principalement de la responsabilité des syndicats nationaux³⁵⁷. En outre, UNI Europa mène également des actions de lobbying et organise des campagnes politiques³⁵⁸.

³⁴⁸ [Global Creators and Performers Demand Creative Rights in AI Proliferation](#), CISAC, 20 juillet 2023.

³⁴⁹ [CISAC VP Ángeles González-Sinde Calls for Ethical Rules on AI Use in the Film Industry](#), CISAC, 20 octobre 2023.

³⁵⁰ Ils organisent par exemple des séminaires, voir [How Will AI Transform the Music Industry? Expert Panel at the IPRS International Music Creators' Seminar Discuss](#), CISAC, 10 janvier 2024.

³⁵¹ [CISAC Annual Report Highlights Its Work Programme on Behalf of CMOs Worldwide](#), CISAC, 22 mai 2024.

³⁵² Ils déclarent que « les créateurs doivent avoir le droit d'accorder des autorisations pour l'utilisation de leurs œuvres par les outils d'IA », Rapport annuel de la CISAC, p. 20, *op. cit.*

³⁵³ La transparence signifie que « les fournisseurs d'IA doivent être obligés d'informer sur l'entraînement à partir d'œuvres protégées par le droit d'auteur », *ibid.*

³⁵⁴ [UNI MEI](#) (International Arts Entertainment Alliance).

³⁵⁵ [Avancer par la négociation collective – Résolutions adoptées par la 5^{ème} conférence d'UNI Europa Bruxelles, 27-29 avril 2021](#), UNI Europa.

³⁵⁶ *Ibid*, page 16.

³⁵⁷ *Ibid*.

³⁵⁸ Voir par exemple, [Le travail numérique dans le secteur des médias, des arts et du divertissement : Défis et Opportunités](#), UNI Global Union Media and Entertainment, FIM, FIA, and EFJ ou [Solidarity with SAG-AFTRA and FIA](#), UNI MEI.



La Fédération internationale des acteurs (FIA) est une fédération mondiale composée de syndicats, de confréries et d'associations professionnelles d'artistes-interprètes³⁵⁹. L'IA est l'une de ses principales préoccupations, comme en témoigne la page qui lui est consacrée sur son site web³⁶⁰. Elle a publié un guide qui explique les implications de l'IA pour les artistes-interprètes que représentent les affiliés de la FIA³⁶¹. Le guide énonce également les principes fondamentaux de la FIA concernant l'utilisation de l'IA tout en offrant des conseils sur « la manière dont les syndicats peuvent structurer leur stratégie de négociation afin de renforcer les protections et les compensations pour leurs membres³⁶² ». Bien que tous les affiliés de la FIA ne soient pas en mesure de négocier des dispositifs de protection vis-à-vis de l'IA dans le cadre de leurs négociations collectives, le guide fournit des conseils aux syndicats qui peuvent le faire, notamment en ce qui concerne le consentement explicite et la rémunération supplémentaire pour le recours aux doubles numériques d'artistes-interprètes³⁶³. Enfin, il propose des pistes d'action sur la manière dont les affiliés de la FIA peuvent activement plaider en faveur de politiques et d'un environnement réglementaire appropriés, tout en soulignant l'importance d'un cadre législatif contraignant³⁶⁴.

La Writers' Guild of Great Britain (WGGB) constitue un autre exemple de syndicat intéressant, qui présente des similitudes avec la WGA. La WGGB est un syndicat qui représente les auteurs professionnels tels que ceux qui travaillent dans les secteurs audiovisuels. Il peut négocier de meilleures conditions de travail avec les principaux acteurs de l'industrie britannique³⁶⁵. Ces accords types avec les grandes organisations profitent à tous les auteurs, mais seuls les membres de la WGGB ont accès à certains avantages³⁶⁶. Contrairement aux États-Unis, où l'adhésion à la WGA est obligatoire dans certains cas, l'adhésion à la WGGB est totalement facultative.

Pour l'instant, la WGGB n'a pas conclu de nouveaux accords contractuels comportant des dispositions relatives à l'utilisation de l'IA, mais elle communique activement sur cette question³⁶⁷. Enfin, elle s'est engagée à soutenir les grèves aux États-Unis, en signe de solidarité avec le « syndicat frère³⁶⁸ ». La WGGB a le pouvoir de négocier de nouveaux accords, comme par le passé, qui pourraient intégrer des dispositions relatives à l'utilisation de l'IA. Il sera intéressant de suivre les actions de la WGGB afin d'observer si cette solution se concrétisera.

³⁵⁹ « [À propos – FIA](#) », FIA.

³⁶⁰ « [Intelligence artificielle – FIA](#) », FIA.

³⁶¹ [Guide de la FIA sur l'intelligence artificielle](#), FIA, 23 novembre 2023.

³⁶² [Principes directeurs et guide pratique de la FIA en matière d'intelligence artificielle](#), FIA, 23 novembre 2023

³⁶³ [Guide de la FIA sur l'intelligence artificielle](#), p. 9 et 10, *op. cit.*

³⁶⁴ *Ibid*, p. 12 et 13.

³⁶⁵ « [About](#) », (Writers' Guild of Great Britain).

³⁶⁶ Jonny Walfisz, « [WGA Strikes: What Is the State of Writers' Unions in Europe?](#) », *Euronews*, 31 mai 2023.

³⁶⁷ Voir par exemple leur déclaration, *Writers and AI* (n° 7).

³⁶⁸ Sarah Woodley, « [WGA Strike](#) », *Writers' Guild of Great Britain*, 2 mai 2023. Voir également [Sarah Woodley](#), « [WGA Strike Ends](#) », *Writers' Guild of Great Britain*, 27 septembre 2023).



6.5. Observations finales : les lacunes à combler et les perspectives d'avenir

Pour l'instant, les discussions en Europe se concentrent surtout sur des questions de propriété intellectuelle telles que le droit d'auteur sur les données utilisées pour entraîner les systèmes d'IA, la rémunération dans le cadre du droit d'auteur, la protection par le droit d'auteur du contenu généré par les systèmes d'IA, la notion de transparence et l'octroi de licences³⁶⁹.

Les discussions et les initiatives portant sur le droit du travail en tant que tel restent assez générales et ne sont pas directement axées sur les défis spécifiques que rencontre le secteur audiovisuel³⁷⁰. La nécessité de sensibiliser et de former les créateurs dans ce domaine est également de plus en plus soulignée, comme en témoignent les rapports et la résolution du Parlement européen³⁷¹.

En conclusion, une piste prometteuse concerne la résolution du Parlement européen, qui témoigne d'une philosophie similaire à celle des Etats-Unis, mais qui n'est pas encore suffisamment détaillée pour être concrète sur le fond. Les différentes actions des diverses parties prenantes révèlent également un besoin et un intérêt croissants pour la réglementation de l'utilisation de l'IA dans le secteur audiovisuel. L'évolution de toutes ces initiatives méritera d'être suivie de près.

³⁶⁹ Voir à ce sujet les chapitres 3 et 4 de la présente étude. Voir également l'article 53 du Règlement sur l'IA et les articles 3 et 4 de la directive (UE) 2019/790.

³⁷⁰ Voir par exemple le Règlement sur l'IA. Voir aussi [Un an après l'arrivée de ChatGPT : Réflexions de l'Obvia sur les enjeux et pistes d'action possibles face à l'IA générative](#), Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA), 15 janvier 2024.

³⁷¹ Voir, par exemple, Parlement européen (2023), recommandation 61, *op. cit.* Voir également l'obligation légale d'assurer une « maîtrise de l'IA » à l'article 4 du Règlement sur l'IA. Voir Commission de l'Intelligence artificielle : IA, *notre ambition pour la France*, mars 2024, recommandant de former les personnes à l'IA.



7. Désinformation et intelligence artificielle dans le secteur audiovisuel

Judit Bayer, Université de Münster

7.1. Définition de la désinformation

Au cours des dernières décennies, la transformation de l'environnement informationnel a fait disparaître les éléments qui structuraient auparavant le domaine de l'information publique, tels que le rôle prépondérant des sources d'information traditionnelles, qui permettaient au public d'évaluer la fiabilité des informations. Les barrières à l'entrée du domaine public sont tombées, l'édition s'est hyper-démocratisée et le contrôle de la qualité des informations publiées a désormais valeur d'exception. Cet énorme potentiel a été exploité non seulement par des contributeurs de bonne foi au débat public, mais aussi par des acteurs malveillants. Ces derniers ont, à titre individuel ou en réseau, produit et diffusé des informations fausses ou manipulatrices pour atteindre leurs objectifs financiers ou politiques en portant préjudice à des individus ou à la société. Au sein de l'espace de communication participatif, un certain nombre de leviers d'amplification tels que, notamment, le partage de messages ou l'attribution de *likes*, favorisent la circulation de la désinformation, qui a souvent un caractère plus sensationnel et plus attractif que l'information réelle. Lorsque des personnes de bonne foi partagent ou publient des informations fausses ou manipulées, ou lorsque l'intention initiale de l'auteur n'est pas connue, on parle de fausse information³⁷². De même, lorsque des fournisseurs notoires de contenus rédactionnels, qui veillent habituellement à garantir la qualité de leurs publications, commettent des erreurs occasionnelles, on peut également parler de fausse information. Le présent article est globalement axé sur la désinformation délibérée, produite et diffusée à l'aide de l'IA, et aborde brièvement les cas de fausse information non intentionnelle liés à la GenAI (intelligence artificielle générative).

³⁷² Wardle C. et Derakhshan H., *Information Disorder*, rapport du Conseil de l'Europe DGI(2017)09.



7.2. Les applications de l'IA dans l'industrie de la désinformation

7.2.1. L'IA générative

Depuis 2022, les modèles de GenAI (IA générative) sont librement accessibles à tous, soit gratuitement, soit moyennant un prix modique. Il s'agit de modèles d'IA à usage général qui sont capables de produire du texte, du son, des images ou des vidéos à partir de requêtes textuelles³⁷³. La facilité d'accès à la technologie GenAI devrait entraîner une augmentation du volume global de contenus de toutes sortes, y compris la désinformation³⁷⁴. Or, à ce jour, le déferlement prévu de désinformation produite par l'IA³⁷⁵ n'a pas eu lieu. Les élections parlementaires européennes de 2024 se sont déroulées de façon relativement calme, en l'absence de campagnes massives de désinformation générées par l'IA, et les élections précédentes semblent également avoir été relativement épargnées à cet égard³⁷⁶.

7.2.2. Textes et images

La différence entre l'amélioration et la manipulation d'un contenu existant et la création d'un nouveau contenu n'est pas toujours évidente. C'est pourquoi le Règlement sur l'IA traite conjointement les contenus générés et les contenus manipulés³⁷⁷. Si l'IA n'est utilisée que pour modifier un contenu existant ou en améliorer la qualité, sans modifier de manière substantielle les données d'entrée, elle est exemptée de l'obligation de transparence. Par ailleurs, si l'éditeur assume la responsabilité éditoriale et que le contenu généré par l'IA fait l'objet d'un processus d'examen humain, il n'est pas nécessaire de signaler la génération ou la manipulation artificielle³⁷⁸.

Néanmoins, l'éthique journalistique invite à la transparence sur l'utilisation de l'IA, et le Règlement européen sur la liberté des médias (EMFA) y fait également référence

³⁷³ Considérant (105) du [Règlement \(UE\) 2024/1689 du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle](#) (le Règlement sur l'IA).

³⁷⁴ Cook L. et Chan K., « [AI could supercharge disinformation and disrupt EU elections, experts warn](#) », AP, 5 juin 2024.

³⁷⁵ Zöldi B., « [Not only images of real suffering, but also AI creations of the Hamas-Israeli war are spreading](#) », *Lakmusz*, 27 novembre 2023. EEAS, [2nd Report on FIMI and Interference Threats](#), janvier 2024.

³⁷⁶ Güttel L. et al., [The European Elections 2024: Between Digital Policies and Radical Right Success Online](#). Oxford Internet Institute, 4 juin 2024 ; Canetta T., « [EU Elections 2024: the battle against disinformation was won, but the attrition war is far from over](#) », EDMO Blog, 11 juin 2024 ; Łabuz M. et Nehring C., « [On the way to deep fake democracy? Deep fakes in election campaigns in 2023](#) », *Eur Polit Sci* (2024).

³⁷⁷ Article 52 1a. du Règlement sur l'IA.

³⁷⁸ Article 52 c. du Règlement sur l'IA.



comme un critère de qualification du « fournisseur de services de médias », qui est tenu de déclarer qu'il ne fournit pas de contenu généré par des systèmes d'intelligence artificielle sans soumettre ces contenus à un réexamen par un être humain ou à un contrôle éditorial³⁷⁹.

Cependant, à mesure que s'estompe la frontière ténue entre amélioration et génération d'images, la démarcation entre images réelles et images « générées » ou « fausses » devient également plus floue. Les informations réelles peuvent être illustrées par des photos générées par l'IA³⁸⁰, tandis que la désinformation stratégique est souvent illustrée par des photos ou des vidéos authentiques mais décontextualisées³⁸¹.

Parmi les modèles d'IA générative, les modèles linguistiques sont actuellement les plus développés et les plus couramment utilisés, notamment pour produire de la désinformation. Paradoxalement, un texte généré par une machine peut sembler plus crédible qu'un texte écrit par un être humain³⁸². Ce phénomène est attribué au « biais de fluidité » et au facteur de « véracité », lorsque les utilisateurs jugent un texte crédible sur la base de critères non pertinents, tels que la qualité du texte. Cela permet à des acteurs malveillants d'exploiter les déductions intuitives des utilisateurs³⁸³. Ces mêmes acteurs peuvent également introduire délibérément des fautes de syntaxe et d'orthographe pour que le texte paraisse plus « organique ».

7.2.3. Hypertrucages (*deep fakes*)

Le terme *deep fakes* désigne des vidéos générées ou manipulées par l'IA. L'obligation de déclaration s'applique de la même manière que pour les autres produits d'IA générative. L'utilisation de l'IA doit également être signalée lorsque le produit est utilisé dans le cadre d'œuvres ou de programmes artistiques, créatifs, satiriques ou de fiction, mais de manière à ne pas entraver l'affichage ou la jouissance de l'œuvre. Les vidéos sont souvent manipulées par un processus de permutation ou d'altération des visages. Le remplacement du visage d'une personne par le visage d'une personnalité publique a souvent été utilisé dans des vidéos de satire politique, à l'instar des caricatures de même ordre³⁸⁴. L'imperfection de la technologie est utilisée comme un atout pour obtenir « l'effet de surprise » nécessaire à la dimension humoristique. Pour une vidéo truquée purement fictive, de nombreux éléments visuels doivent être générés par l'IA, tels que l'arrière-plan, les effets de lumière et les mouvements. De tels trucages ne sont pas

³⁷⁹ Article 18 1 e) du Règlement européen sur la liberté des médias.

³⁸⁰ Voir Ghost Archive, images d'archives de la guerre à Gaza, <https://ghostarchive.org/archive/aFK4n> ; Fingas J., « [Adobe Accepts AI-Generated Stock Art, with Limits: The Company Thinks It Can Minimize the Risk of Copyright Disputes](#) », *Engadget*, 5 décembre 2022.

³⁸¹ <https://rtl.hu/belfold/2023/11/23/orosz-propaganda-dezinformacio-magyar-kormanykozeli-sajto> .

³⁸² Zellers R. et al., *Defending against neural fake news*. *Advances in neural information processing systems* 32, 2019.

³⁸³ Hanson R., Grissom A. R. et Mouton C. A., *The Future of Indo-Pacific Information Warfare: Challenges and Prospects from the Rise of AI*, RAND Corporation, 2024, p. 4.

³⁸⁴ RoW, « [A political speech meets a cinematic betrayal](#) », Rest of World 2024, AI Elections Tracker, 2024.



faciles à réaliser : on peut en voir dans des films, mais pas (encore) dans les contenus de désinformation gratuits en ligne. Une autre technique très répandue consiste à modifier le visage, par exemple avec la synchronisation labiale. Des outils simples disponibles en ligne permettent de modifier des vidéos existantes de personnalités publiques de manière à faire dire à la personne autre chose que ce qu'elle dit réellement, comme avec les hypertrucages ciblant, entre autres, Maria Ressa³⁸⁵, un quotidien allemand³⁸⁶ et plusieurs autres victimes à des fins criminelles et pour des intérêts privés.

Ces hypertrucages ne sont pas encore assez répandus pour avoir un impact significatif sur le débat public. Toutefois, en s'immisçant dans des débats présentant des enjeux majeurs, ils comportent un risque spécifique d'induire des décisions individuelles ou politiques importantes sur la base de fausses impressions, comme par exemple l'hypertrucage montrant le général Valery Zaluzhny, commandant en chef des forces armées d'Ukraine, annoncer que le président Zelensky avait assassiné son adjoint et se rendait, tout en appelant les Ukrainiens à faire un coup d'État³⁸⁷. Ces hypertrucages peuvent également interférer dans les relations privées et professionnelles. Par exemple, lorsqu'une visioconférence sur Zoom est interrompue par « Elon Musk »³⁸⁸, cela peut être considéré par certains comme un piratage amusant, mais c'est également le signe d'une fuite potentielle de secrets d'affaires et d'une tromperie en temps réel de hauts fonctionnaires. À l'instar de la campagne de désinformation *Doppelgänger*, cela pose la question de la cybersécurité³⁸⁹. L'opération de désinformation russe soigneusement coordonnée, baptisée *Doppelgänger* par l'organisation EU Disinfo Lab, a cloné au moins 17 sites internet d'authentiques médias européens aux fins de tromper les utilisateurs des sites de ces médias³⁹⁰.

7.2.4. Contenus audio

Selon l'EDMO, les sons générés par l'IA constituent actuellement la plus grande préoccupation en matière de désinformation³⁹¹. Cette technologie est pratiquement aussi développée que la production de textes et, conjointement, ces deux procédés peuvent amplifier mutuellement leur potentiel et leur impact global. La production de contenus

³⁸⁵ Vera Files, « [VERA FILES FACT CHECK: Maria Ressa NOT endorsing 'bitcoin platform](#) », *VERA Files*, 28 février 2024.

³⁸⁶ Reveland C. et Siggelkow P., « [Falsche tagesschau-Audiodateien im Umlauf](#) », *Tagesschau*, 13 novembre 2023.

³⁸⁷ EEAS, [2nd Report on FIMI and Interference Threats](#), janvier 2024.

³⁸⁸ Fanatical Futurist by 311 Institute, [Deep fake Elon Must bombs a Zoom call](#), Youtube, 2021.

³⁸⁹ Alaphilippe A. et al., [DoppelGanger: Media Clones Serving Russian Propaganda](#), EU Disinfo Lab, 27 septembre 2022 ; Goujard C., « [Big, bold and unchecked: Russian influence operation thrives on Facebook](#) », *Politico*, avril 2024, Cook L. et Chan K., « [AI could supercharge disinformation and disrupt EU elections, experts warn](#) », *AP*, 5 juin 2024.

³⁹⁰ Conseil de l'UE, [Manipulation de l'information dans la guerre d'agression de la Russie contre l'Ukraine : l'UE inscrit sept personnes et cinq entités](#), communiqué de presse du 28 juillet 2023 ; Bouchaud P., Faddoul M. et Buse Cetin R., [No Embargo in sight](#), *AI Forensics*, 2024.

³⁹¹ EDMO, [Prebunking AI-generated disinformation ahead of EU elections](#), 24 mars 2024.



audio est « facile et bon marché »³⁹² et plusieurs tests ont permis de générer de fausses déclarations électorales très convaincantes³⁹³. Cela tient du miracle qu'il n'y ait pas eu davantage d'incidents connus. Seules quelques opérations de désinformation politique par imitation vocale ont été signalées, aux États-Unis, au Royaume-Uni³⁹⁴, en Slovaquie et au Soudan. En Slovaquie, ce procédé a été utilisé de façon stratégique au cours des derniers jours de la campagne législative de 2023, empêchant ainsi les médias traditionnels de discuter et de révéler la « fuite » audio en ligne d'un candidat qui aurait prétendument planifié une fraude électorale³⁹⁵. Dans l'affaire soudanaise, le problème ne se situe pas au niveau du contenu audio, qui provient d'un blogueur de bonne foi, mais de la pseudo-réapparition publique de l'ancien dirigeant du Soudan, accusé de crimes de guerre, qui avait disparu de la scène publique depuis un an et que l'on pensait gravement malade³⁹⁶.

7.2.5. Robots

Si la GenAI est le phénomène dont on parle le plus, il existe néanmoins d'autres types de robots qui sont extrêmement actifs et fonctionnels dans la diffusion de la désinformation. Ces robots automatisent la diffusion, simulent la participation aux discussions et le soutien de certains messages. Associés à la GenAI, ils peuvent concevoir et mener automatiquement une campagne de désinformation en ligne, en distribuant automatiquement des messages personnalisés basés sur le profil comportemental des utilisateurs sans qu'il soit nécessaire de faire intervenir des « *trolls* » humains.

Un exemple intéressant nous est fourni par « Grok », le *chatbot* de X, qui est censé donner un contexte et un titre à des thèmes d'actualité et qui a falsifié une vidéo virale montrant une attaque de missiles lors de la guerre en Ukraine. Il a amplifié la propagation de cette vidéo avec un titre mentionnant une attaque de l'Iran contre Tel Aviv³⁹⁷. Bien que cet incident ne s'inscrive pas dans une campagne de désinformation coordonnée, la diffusion de fausses informations par des robots peut avoir un effet similaire. Les robots sont également capables de déployer des stratégies relevant de l'anthropomorphisme, telles que la publication de messages à des intervalles irréguliers et parfois dénués de cohérence, la production délibérée de fautes d'orthographe ou l'utilisation de mots et d'expressions à la mode. Parallèlement, on assiste à une campagne intensive de développement et de tests de *chatbots* empathiques dans le secteur de la santé et de l'éducation, voire même du service à la clientèle. Même si les avis des experts divergent

³⁹² Jingnan H., « [It's quick and easy to clone famous politicians' voices, despite safeguards](#) », *NPR*, Untangling Disinformation, 31 mai 2024.

³⁹³ Swenson A., « [Tests find AI tools readily create election lies from the voices of well-known political leaders](#) », *AP News*, 31 mai 2024.

³⁹⁴ Meaker M., « [Deep fake Audio is a Political Nightmare](#) », *Wired*, 9 octobre 2023.

³⁹⁵ Kóváry K., « [Slovakia: Deep fake audio of Denník N journalist offers worrying example of AI abuse](#) », *IPI NEWSROOM*, 31 octobre 2023.

³⁹⁶ Goodman J. et Hashim M., « [AI: Voice cloning tech emerges in Sudan civil war](#) », *BBC*, 5 octobre 2023.

³⁹⁷ Thomas J., « [No, Iran has not started attacking Tel Aviv](#) », *Euronews*, 11 avril 2024.



en la matière³⁹⁸, l'objectif pourrait être plus facile à atteindre qu'on ne le pense : des tests ont montré que le *chatbot* de Google était plus empathique que de vrais médecins³⁹⁹. Dans le sillage des expériences ainsi menées, les groupes extrémistes utilisent également les *chatbots* pour recruter et convaincre leurs partisans. Des chercheurs ont établi que des robots extrémistes étaient mieux à même de persuader les sympathisants d'adhérer que des recruteurs humains. Ils estiment que cela est dû à la tonalité optimiste du robot, à la rapidité de ses réponses et à son absence de jugement à l'égard des sympathisants⁴⁰⁰.

7.2.6. IA et fausses informations

L'entraînement des modèles de GenAI à l'aide de données non fiables, notamment de la désinformation, conduit à la production de fausses informations. En dépit de leur apparente précision, les modèles ne sont pas capables de vérifier eux-mêmes la véracité des faits.

L'entraînement de l'IA n'a pas pour objectif de dispenser la vérité, ni de soumettre le thème de la discussion à une évaluation critique, mais tout simplement d'agrèger et reproduire un contenu déjà existant. Cela conduit à des fabulations occasionnelles (communément appelées « hallucinations »)⁴⁰¹. Les fausses informations et les fabulations produites par les LLM (grands modèles de langages) risquent alors d'être utilisées et diffusées par erreur dans le cadre d'un produit journalistique, polluant ainsi davantage l'océan informationnel.

Distinguer le vrai du faux est un véritable défi philosophique, sémantique, juridique et épistémologique. On ne saurait donc confier cette tâche à des *chatbots* d'IA, même à long terme. Compte tenu de l'ampleur que pourraient prendre les agents conversationnels, il est légitime de s'inquiéter et de prévoir la mise en place d'un cadre réglementaire. Les *chatbots* vocaux peuvent donner l'illusion parfaite d'un interlocuteur intelligent et sont déjà intégrés dans les smartphones.

³⁹⁸ Seitz L., « [Artificial empathy in healthcare chatbots: Does it feel authentic?](#) », *Computers in Human Behavior: Artificial Humans*, vol.2, numéro 1, 2024.

³⁹⁹ Quach K., « [Google AI chatbot more empathetic than real doctors in tests](#) », *The Register*, 16 janvier 2024.

⁴⁰⁰ Mantello P. et Ho M. T., « [Losing the information war to adversarial AI](#) », *AI & Society*, 28 avril 2023.

⁴⁰¹ Le terme « hallucination » est impropre, car l'hallucination désigne une fausse perception sensorielle, ce qui n'est pas le cas de la GenAI. Des scientifiques ont proposé le mot « fabulation », qui désigne une « reconstruction erronée d'informations influencée par les connaissances, les expériences, les attentes et le contexte existants », Smith A. L., Greaves F. et Panch T., « [Hallucination or confabulation? Neuroanatomy as metaphor in large language models](#) », *PLOS Digital Health* 1^{er} novembre 2023 ; Maleki N., Padmanabhan B. et Dutta, K., « [AI Hallucinations: A Misnomer Worth Clarifying](#) », *arXiv preprint arXiv:2401.06796* (2024) ; « [Emsley, R. ChatGPT: these are not hallucinations – they're fabrications and falsifications](#) », *Schizophr* 9, 52, 2023.



7.3. La lutte contre la désinformation

7.3.1. Réglementation

7.3.1.1. Dispositions législatives

La qualité du contenu ne pouvant servir de base à une interdiction, les outils juridiques se concentrent sur les techniques qui amplifient l'impact de la tromperie et de la manipulation. Comme indiqué ci-dessus, le Règlement sur l'IA exige que l'utilisation d'hypertrucages et d'IA générative soit rendue transparente. Plus précisément, les fournisseurs de systèmes d'IA générique doivent veiller à ce que les contenus produits par leurs systèmes soient marqués d'un filigrane qui permette de les identifier⁴⁰². Les techniques délibérément manipulatrices ou trompeuses sont également interdites, mais seulement dans certaines conditions spécifiques. Parmi les motifs justifiant une interdiction légale, on note une utilisation desdites techniques ayant pour but ou pour effet d'altérer substantiellement le comportement d'une personne, d'entraver la capacité de la personne manipulée ou trompée à prendre une décision éclairée, et le fait que la décision prise cause ou soit susceptible de causer un préjudice important aux personnes concernées⁴⁰³. À titre d'exemple de scénario répondant à cette définition, citons les arnaques téléphoniques où l'appelant imite la voix d'un proche de la victime pour lui demander de transférer une somme d'argent⁴⁰⁴.

Parmi les risques systémiques des très grandes plateformes en ligne et des très grands moteurs de recherche (ci-après dénommés conjointement VLOPSE), la DSA énumère, entre autres, la diffusion de contenus illicites et « tout effet négatif réel et prévisible sur le discours civique, les processus électoraux et la sécurité publique. » La lutte contre les contenus illicites se situe donc à deux niveaux : d'une part, ces contenus doivent être retirés lorsque le fournisseur en a connaissance, et d'autre part, ils sont considérés comme un risque systémique. Cette dernière disposition s'appuie sur l'obligation fondamentale de veiller à garantir une plateforme sécurisée. Lorsqu'elles procèdent à de telles évaluations des risques, les VLOPSE doivent notamment prendre en compte leurs systèmes algorithmiques et de recommandation, ainsi que leurs systèmes de modération des contenus⁴⁰⁵. Dès lors qu'un fournisseur de services constate une manipulation intentionnelle de son service ou des contenus commerciaux trompeurs et de grande diffusion, tels que l'utilisation non authentique du service par des robots, de faux comptes ou de vastes programmes de désinformation, il est tenu de retirer le contenu ou d'en limiter la visibilité. La dernière option fait intervenir des mesures telles que la rétrogradation du contenu dans les systèmes de classement, le blocage de l'utilisateur à

⁴⁰² Article 52 1a du Règlement sur l'IA.

⁴⁰³ Article 5 du Règlement sur l'IA.

⁴⁰⁴ Bethea C., « [The Terrifying A.I. Scam That Uses Your Loved One's Voice](#) », *The New Yorker*, 7 mars 2024.

⁴⁰⁵ Article 34 de la DSA.



son insu (bannissement silencieux) ou la démonétisation du service. Il n'y a pas lieu de fournir un exposé des motifs lorsque la mesure sanctionne une utilisation non authentique du service, néanmoins l'utilisateur doit toujours disposer d'un droit de recours effectif devant les juridictions nationales⁴⁰⁶.

7.3.1.2. Le Code de bonnes pratiques contre la désinformation

La DSA est complétée par des codes de bonnes pratiques, notamment le Code de bonnes pratiques renforcé contre la désinformation (2022)⁴⁰⁷ qui est, à cet égard, le plus pertinent⁴⁰⁸. Plutôt que de s'attaquer directement au contenu de la désinformation, ce code s'attache à définir les circonstances spécifiques de la communication en ligne, en distinguant notamment deux catégories : la publicité et les pratiques manipulatoires. Les signataires du code s'engagent à séparer la publicité de la désinformation sur l'ensemble de la chaîne de valeur de la publicité, y compris les services de paiement électronique en ligne, les plateformes de commerce électronique, les systèmes de financement participatif (*crowdfunding*) ou de donation. Le code intègre également le projet de règlement de l'UE relatif à la transparence et au ciblage de la publicité à caractère politique⁴⁰⁹ afin de garantir la transparence avant même son entrée en vigueur⁴¹⁰.

En ce qui concerne les pratiques manipulatoires assistées par l'IA, le code mentionne spécifiquement les faux comptes, la prise de contrôle de comptes (comme la campagne *Doppelgänger*), l'amplification par des robots, le piratage et la fuite de données, l'usurpation d'identité et les hypertrucages malveillants⁴¹¹. Le code invite les développeurs ou les opérateurs de systèmes d'IA qui sont en mesure de créer ou de diffuser des hypertrucages et d'autres contenus générés par l'IA à s'engager à respecter les principes de transparence préconisés et l'interdiction des pratiques de manipulation spécifiées dans le Règlement sur l'IA. Par ailleurs, le code comporte certains engagements visant à responsabiliser les utilisateurs par le développement de l'éducation aux médias, de la sécurité et de la responsabilité dans la conception des plateformes, et à responsabiliser la

⁴⁰⁶ Considérant 55 de la DSA.

⁴⁰⁷ [Code de bonnes pratiques renforcé contre la désinformation](#), 16 juin 2022. Parmi les autres codes traitant de thématiques connexes, citons le Code de conduite contre les discours haineux illicites en ligne, les codes de conduite pour la publicité en ligne (article 46 de la DSA) et les codes de conduite pour l'accessibilité (article 47 de la DSA).

⁴⁰⁸ Pour de plus amples informations sur le Code de bonnes pratiques contre la désinformation, voir le rapport de l'Observatoire européen de l'audiovisuel : Cabrera Blázquez F.J., Cappello M., Talavera Milla J., Valais S., [La responsabilisation des utilisateurs face à la désinformation en ligne](#), IRIS Plus, Observatoire européen de l'audiovisuel, Strasbourg, 2022.

⁴⁰⁹ Ce règlement a été publié le 20 mars 2024, mais il n'entrera en vigueur que 18 mois après sa publication au Journal officiel. Grieria M., « [EU agrees on political advertising rulebook effective after Parliament elections](#) », *Euractive*, 7 novembre 2023.

⁴¹⁰ Engagements 4 à 9, *Strengthened Code of Practice on Disinformation*.

⁴¹¹ Les pratiques d'« achat de faux engagements » et « d'amplification artificielle de l'audience ou du soutien public perçu à l'égard de la désinformation » étant, à mon avis, incluses dans « l'amplification par les robots », elles ne sont donc pas mentionnées spécifiquement dans la liste. Voir l'engagement 14.



communauté des vérificateurs de faits, élargissant ainsi le champ d'application de l'obligation d'éducation à l'IA visée par le Règlement sur l'IA ⁴¹².

7.3.1.3. Application de la législation

Au moment de la rédaction du présent rapport, les très grandes plateformes ont publié leur troisième rapport de transparence⁴¹³. La plupart d'entre elles ont déployé des efforts appréciables pour lutter contre la désinformation et se prémunir contre les systèmes d'IA générative sur leurs services⁴¹⁴. Cependant, en dépit de leurs efforts pour se conformer aux recommandations du code et concevoir des engagements mesurables, une analyse récente établit que le niveau de conformité n'est pas à la hauteur des attentes⁴¹⁵. Au-delà des doutes quant à la conformité avec la DSA de leurs politiques et pratiques de lutte contre la publicité trompeuse et de transparence à l'égard des contenus politiques, la procédure épingle également l'absence d'outils efficaces de surveillance de l'IA avant les élections européennes et la suppression de CrowdTangle⁴¹⁶ sans qu'un dispositif équivalent ne vienne le remplacer⁴¹⁷. La communauté des chercheurs et des activistes proteste contre la mise hors service de cet outil⁴¹⁸.

Le Parlement allemand a rédigé un projet de loi prévoyant une pénalisation des hypertrucages réalisés en violation des droits de la personnalité d'un tiers (sans son consentement)⁴¹⁹. Une telle révision du droit pénal rendrait les hypertrucages illégaux en Allemagne et donc soumis à l'obligation de retrait prévue par la DSA. Etant donné que cette dernière renvoie la définition du terme « illégal » à la compétence des Etats membres, d'autres Etats pourraient également suivre cette voie.

⁴¹² Législation européenne sur l'IA, [Article 4 : Maîtrise de l'IA](#).

⁴¹³ Commission européenne, [Online platforms put special focus on elections in the third batch of reports under the Code of Practice on Disinformation](#), Affaires publiques Bruxelles, 26 mars 2024.

⁴¹⁴ EDMO, [Prebunking AI-generated disinformation ahead of EU elections](#), 24 mars 2024.

⁴¹⁵ EDMO News, [EFCSN review of the fulfillment of the Code of Practice on Disinformation by the very large online platforms and search engines](#), 11 janvier 2024 ; Hernández-Echevarría C., [Major Tech Platforms Fail to Deliver on EU Fact-Checking Commitments, Risking DSA Compliance](#), 11 janvier 2024 ; Lai S. et Yadav K., « [Operational Reporting in Practice: The EU's Code of Practice on Disinformation](#) », *Carnegie Endowment for International Peace*, 21 novembre 2023.

⁴¹⁶ CrowdTangle est un outil d'analyse de contenu développé par Facebook qui permet d'analyser toutes les publications publiques diffusées par Meta, y compris les données d'engagement, telles que les partages, les vues, les commentaires, les *likes* et autres réactions. Il était utilisé par des éditeurs, des journalistes, des chercheurs et des vérificateurs de faits. Voir Tess, [What data is CrowdTangle tracking?](#), Crowdangle, 2023-2024.

⁴¹⁷ Albert J., « [Facebook's gutting of CrowdTangle: a step backward for platform transparency](#) », *Algorithmwatch*, 3 août 2022.

⁴¹⁸ [Letter Urging Meta to Maintain CrowdTangle Tool Through Upcoming Elections](#), 15 mai 2024. Voir aussi [Open Letter to Meta](#), Mozilla Foundation, 2024.

⁴¹⁹ Projet de loi du Bundestag, [Entwurf eines Gesetzes zum strafrechtlichen Schutz von Persönlichkeitsrechten vor Deep fakes](#) 14 mai 2024.



7.3.2. Vérification des faits avec l'aide de l'IA

La vérification des faits permet de clarifier ce qui se passe dans l'univers informationnel et de réfuter des éléments de désinformation significatifs. Même si le fait de riposter à chaque message de désinformation peut sembler vain et stérile, il est important de connaître les tendances à l'œuvre pour pouvoir élaborer des stratégies appropriées.

7.3.2.1. Repérer la désinformation avec les méthodes traditionnelles

Même si la GenAI est en mesure de renforcer la crédibilité des contenus de désinformation produits, les circuits de distribution des fausses informations restent identiques, de sorte que les outils de vérification des faits sont restés opérationnels⁴²⁰. Toutefois, l'évaluation qualitative des contenus nécessite toujours une intervention humaine⁴²¹.

Z. Guo *et al.* ont défini trois étapes pour identifier un contenu de désinformation : 1) détection d'une allégation, 2) recherche de preuves et 3) vérification de l'allégation⁴²². Nourrir des soupçons sur la seule base du contenu nécessite une formation intensive sur une grande quantité de données et ce n'est pas, actuellement, la méthode la plus prometteuse⁴²³. En revanche, la méthode qui s'avère la plus efficace consiste à combiner la reconnaissance de modèles suspects de diffusion, de propagation et de type de locuteur⁴²⁴. L'IA peut également être utile dans la recherche de preuves, par exemple pour trouver des informations contextuelles supplémentaires par rapport à l'allégation initiale. La vérification de l'allégation est la tâche la plus complexe : elle comprend l'interprétation et l'évaluation des informations, ainsi que leur comparaison avec l'ensemble des preuves ayant pu être établies⁴²⁵. Les technologies actuelles de vérification artificielle des faits sont mal adaptées pour mener à bien cette tâche, mais des expériences sont actuellement en cours avec des LLM⁴²⁶. Toutefois, le problème majeur concerne la fiabilité des informations pouvant servir de preuve. Il est nécessaire de mettre en place une supervision manuelle par du personnel humain⁴²⁷ ou une base de données préparée par des humains et contenant des informations préalablement vérifiées. En

⁴²⁰ EEAS, *2nd Report on FIMI and Interference Threats*, janvier 2024.

⁴²¹ EDMO, *Fact-Checking Overview*.

⁴²² Guo Z., Schlichtkrull M. et Vlachos A., « *A Survey on Automated Fact-Checking* », *Transactions of the Association for Computational Linguistics 2022*, 10 178-206.

⁴²³ Pour plus d'informations, voir Grimme, Ch. *et al.* (éd.), *Disinformation in Open Online Media: First Multidisciplinary International Symposium*, MISDOOM 2019, Hambourg, Allemagne, 27 février – 1^{er} mars 2019, Revised Selected Papers, vol 12021 (Springer International Publishing 2020).

⁴²⁴ McGovern, A., « *Artificial Intelligence System Could Help Counter the Spread of Disinformation* », *MIT News | Massachusetts Institute of Technology*. Voir également Smith S. T. *et al.*, *Automatic detection of influential actors in disinformation networks*, *Computer Sciences*, 7 janvier 2021 ; voir aussi Cassauwers T., « *Can Artificial Intelligence Help End Fake News?* », *Research and Innovation*, 15 avril 2019.

⁴²⁵ Sittmann J., Tompkins A., « *The strengths and weaknesses of automated fact-checking tools* », *Deutsche Welle Akademie*, 17 juillet 2020.

⁴²⁶ Guo Z., Schlichtkrull M. et Vlachos A., « *A Survey on Automated Fact-Checking* », *Transactions of the Association for Computational Linguistics 2022*, 10: 178-206.

⁴²⁷ Borel B., *The Chicago guide to fact-checking*, University of Chicago Press, 2023.



particulier dans les cas où l'information n'est pas facilement disponible, par exemple en situation de crise ou en raison de restrictions délibérées de la part d'un État autoritaire, la gestion d'une base de données actualisée et exhaustive constitue un défi considérable, voire insurmontable⁴²⁸. Les vérificateurs de faits humains peuvent toutefois faire preuve d'inventivité pour découvrir des indices cachés permettant de réfuter des allégations, par exemple en présence de vidéos décontextualisées.

Une fois les données disponibles, elles peuvent être structurées par l'IA pour les rendre facilement accessibles à d'autres outils d'IA en tant que bases de données fiables pour la vérification automatisée des faits (par exemple, ClaimReview)⁴²⁹, réduisant ainsi les thèmes propices à la désinformation. En outre, l'IA permet également de signaler les documents déjà vérifiés, de les diffuser et de rationaliser la collaboration internationale. L'IA est également utilisée pour simplifier le processus de publication (Chequedo) ou pour partager les résultats de la vérification des faits en temps réel par le biais d'une application mobile, y compris en direct pendant les discours politiques (Factstream)⁴³⁰.

La reconnaissance des récits de propagande à l'aide de l'analyse de l'IA s'avère également fructueuse. Dans une étude, un système d'apprentissage machine a été entraîné à enregistrer la fréquence de séries complètes de thèmes d'actualité, plutôt que des termes individuels, dans les médias grand public d'un pays afin de créer une base de référence. L'application a ensuite associé chaque mot à un vecteur multidimensionnel afin d'établir une relation entre les termes, sur la base des angles et distances des vecteurs, à l'aide de quelques centaines de paramètres dimensionnels. Ensuite, le modèle s'est vu attribuer des termes antagonistes (bon-mauvais) en fonction desquels chaque article a été catégorisé sur la base de la matrice du texte analysé par vecteur, avec la possibilité de couvrir plusieurs phrases plutôt que des termes isolés. Les résultats de l'IA montrent qu'en septembre 2021, la propagande pro-russe dans la presse grand public hongroise (d'obédience gouvernementale) s'est considérablement accrue. Les travaux indépendants menés par des chercheurs humains parviennent aux mêmes résultats. Le modèle peut être constamment mis à jour et répond à des demandes similaires en quelques minutes⁴³¹. Ce modèle applique le procédé bien établi d'apprentissage semi-supervisé, qui combine un volume raisonnable de travail humain avec des données fiables et permet d'obtenir ainsi les meilleurs résultats⁴³². Le développement du modèle a requis de la main d'œuvre et il est resté lié au contexte, mais le modèle est capable de reconnaître des informations plus complexes que les modèles précédents.

⁴²⁸ Graves L., *Understanding the Promise and Limits of Automated Fact-Checking*, Reuters Institute, Oxford, février 2018.

⁴²⁹ Voir le site de ClaimReview : <https://www.claimreviewproject.com/>

⁴³⁰ Adair B., « [FactStream app now shows latest fact-checks from Post, FactCheck.org and PolitiFact](#) », *Reporterslab*, 7 octobre 2018.

⁴³¹ Ries U., « [AI Helps Uncover Russian State-Sponsored Disinformation in Hungary](#) », *DarkReading*, 20 novembre, 2023. Voir en particulier l'étude citée menée par Novak et Weddingensen.

⁴³² Xin, L. and others, « [A Novel Self-Learning Semi-Supervised Deep Learning Network to Detect Fake News on Social Media](#) », *Multimedia Tools and Applications*, 2021.



7.3.2.2. Reconnaissance de la GenAI par l'IA

Il arrive que certains textes rédigés par l'IA paraissent plus crédibles que des écrits humains, en particulier lorsque des efforts sont déployés pour atténuer la perfection des produits synthétiques⁴³³. Cependant, les développeurs mènent des travaux ciblés pour pouvoir identifier les contenus générés par l'IA. Par exemple, les outils de détection d'images synthétiques sont axés sur les propriétés des architectures génératives⁴³⁴. D'autres chercheurs ont constaté que la représentation d'une personne sur des vidéos ou des enregistrements audio manipulés ne correspond pas, dans les moindres détails, à la personne réelle, ce qui peut être détecté par l'IA à condition de disposer d'une vidéo authentique montrant le visage du locuteur ou de la locutrice, ou d'un enregistrement audio authentique⁴³⁵. Néanmoins, on ne peut pas appliquer la même logique au texte, car le texte est construit en tant que tel ; en outre, de nombreux textes authentiques sont repris et améliorés par l'IA. Même si le Règlement sur l'IA prévoit l'obligation de placer un filigrane, celui-ci est, semble-t-il, facile à supprimer. De même, il est facile d'ajouter un faux filigrane pour discréditer un contenu authentique, ou encore d'ajouter un filigrane falsifié censé prouver que l'image provient d'un véritable appareil photo⁴³⁶. En conclusion, les outils d'IA sont capables d'aider les humains éclairés à détecter la désinformation et à extrapoler les résultats pour atteindre des publics plus larges, mais il convient de ne pas négliger le rôle de l'humain dans la boucle⁴³⁷. De plus, ils peuvent être utiles pour produire et diffuser des informations facilement accessibles et véridiques, avec un « filigrane » de confiance.

7.4. Conclusion

L'un des plus grands défis posés par la GenAI réside dans le volume et la rapidité de production des contenus et dans la capacité illimitée d'adapter les messages à divers publics, renforçant ainsi la crédibilité de la désinformation par la personnalisation⁴³⁸.

Le risque majeur lié au potentiel de production de contenus par l'IA est, une fois de plus, la perte d'une nouvelle part de nos ancrages de confiance : niveau de qualité supérieure,

⁴³³ Zellers R. *et al.*, *Defending against neural fake news*, *Advances in neural information processing systems* 32 (2019).

⁴³⁴ R. Corvi *et al.*, *On The Detection of Synthetic Images Generated by Diffusion Models*, ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Île de Rhodes, Grèce, 2023, p. 1-5.

⁴³⁵ Cozzolino D. *et al.*, *Audio-visual person-of-interest deep fake detection*, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023 (p. 943-952).

⁴³⁶ Hoffman-Andrews J., « [AI Watermarking Won't Curb Disinformation](#) », *EFF*, 5 janvier 2024.

⁴³⁷ Montoro-Montarroso A. *et al.*, « [Fighting disinformation with artificial intelligence: fundamentals, advances and challenges](#) », *Network activism*, vol. 32, n° 3, 2023.

⁴³⁸ Wach K. *et al.*, « [The dark side of generative artificial intelligence: controversies and risks of ChatGPT](#) », *Entrepreneurial Business and Economics Review* 11, n° 2, 2023, 7-24. Voir aussi Dobber T. *et al.*, « [Do \(Microtargeted\) Deep fakes Have Real Effects on Political Attitudes?](#) », *International Journal of Press/Politics* 26, n° 1, 2021, p. 69-91.



bonne structuration, choix des mots ou valeur probante de l'image et de la vidéo. Les manifestations de la réalité de notre épistémie sont de plus en plus troubles. Il est indispensable de développer de nouveaux ancrages de confiance et systèmes de preuves.

Il convient également de se préparer à traiter l'autre aspect du problème, qui consiste à jeter le discrédit sur un contenu authentique en alléguant qu'il a été généré par l'IA (par exemple en affirmant que la photo officielle du Premier ministre bulgare tenant un discours devant le Parlement européen est un faux⁴³⁹). Les faux positifs des systèmes de vérification des faits librement accessibles peuvent alimenter la méfiance. L'exploitation du battage médiatique autour de la menace que représente l'IA⁴⁴⁰ est également une arme qui peut engendrer une défiance générale à l'égard du système démocratique, comme l'affirmation que tout contenu peut être manipulé et que, par conséquent, la vérité objective n'existe pas⁴⁴¹.

⁴³⁹ Bontcheva K., *Generative AI and Disinformation : Recent Advances, Challenges, and Opportunities*, Livre blanc, projet Vera.ai, février 2024.

⁴⁴⁰ EEAS, *2nd Report on FIMI and Interference Threats*, janvier 2024.

⁴⁴¹ Political Capital, « A significant part of Hungarians already doubt the existence of facts and reality », *PCBlog.Átlátszó*, 30 novembre 2023.



8. Pluralisme et diversité

Mira Burri, Université de Lucerne

Le présent chapitre se penche sur les opportunités offertes par l'IA et se propose d'analyser la nature et la portée de leur impact sur l'exercice de la liberté d'expression, ainsi que sur le pluralisme des médias et la diversité culturelle. Tout en reconnaissant certains effets positifs, tels que l'amélioration potentielle de l'accessibilité et de l'efficacité dans la création des contenus, ce chapitre souligne en particulier les effets négatifs de l'IA, tels que la personnalisation des contenus, les représentations tendancieuses, le filtrage des contenus et le contrôle d'accès, ainsi que, à plus grande échelle, l'impact préjudiciable qu'elle exerce sur les modèles de médias traditionnels en sapant la fiabilité de l'information disponible, en faussant le débat public et en entravant la diversité culturelle.

8.1. Contexte général : l'IA, une technologie disruptive

La transition numérique liée à l'avènement et la généralisation de l'Internet ne cesse de modifier les modalités de production, de diffusion, d'accessibilité et de consommation des contenus ; elle a également transformé les modèles en matière d'expérience utilisateur et de participation des utilisateurs, avec des répercussions négatives sur le pluralisme et la diversité⁴⁴². En tant que processus porté par de nouveaux développements technologiques, l'IA marque une nouvelle étape qui ne manquera pas de se traduire par une mutation disruptive et l'émergence d'un certain nombre de défis spécifiques. Ce chapitre se propose d'explorer les opportunités liées à l'IA, c'est-à-dire ses capacités actuelles et potentielles, et son impact sur la liberté d'expression, le pluralisme des médias et la diversité culturelle. Il expose les préoccupations spécifiques qui ont été soulevées à cet égard et qui pourraient nécessiter des mesures d'ordre réglementaire ou autre pour faire face à ce nouvel environnement afin de préserver les valeurs fondamentales de la société tout en profitant des bénéfices de l'IA dans un univers culturel pluraliste, sain et durable.

⁴⁴² Voir par ex. Y. Benkler, *The Wealth of Networks: How social production Transforms Markets and Freedom*, Yale University Press, 2006) ; C. R. Sunstein, *Republic.com 2.0*, Princeton University Press, 2007.



En premier lieu, il convient de noter qu'il n'existe pas de définition communément admise de l'IA – non seulement parce que la technologie évolue très rapidement, mais aussi parce que les perceptions, les contextes sociétaux et les sensibilités éthiques peuvent varier. Pour les besoins de ce chapitre, nous utiliserons la définition (récemment mise à jour) de l'Organisation de coopération et de développement économiques (OCDE), qui énonce de manière neutre et succincte les différentes opportunités liées à l'IA⁴⁴³. Alors que l'impact sur le secteur culturel des systèmes d'IA conventionnels, qui servent essentiellement à analyser des données et faire des prévisions, a déjà été débattu, sans pour autant aboutir à des résultats concluants⁴⁴⁴, l'intelligence artificielle générative (GenAI), plus récente, pose de nouveaux défis. En particulier si l'on considère que ces systèmes d'IA, tels que ChatGPT, peuvent générer de *nouveaux* contenus et produits qui imitent, tout en pouvant également les dépasser, les modèles et structures présents dans leurs données d'apprentissage⁴⁴⁵.

L'IA est généralement reconnue comme une technologie à usage général (*general purpose technology* - GPT)⁴⁴⁶ dotée d'un énorme potentiel de transformation dont les répercussions sur les différents secteurs et industries dépassent ses applications initiales⁴⁴⁷. En outre, et peut-être à la différence d'autres technologies à usage général, telles que l'imprimerie, l'électricité ou encore Internet, l'IA est de nature itérative – au sens où les systèmes d'IA apprennent et adaptent continuellement leurs performances en les exposant à de nouvelles données, expériences et boucles de rétroaction⁴⁴⁸. Les GPT en général et l'IA en particulier posent un défi majeur en matière de réglementation, car leur évolution et leur diffusion ne sont ni linéaires, ni prévisibles⁴⁴⁹.

⁴⁴³ La définition de l'OCDE et la définition du règlement de l'UE sur l'IA sont présentées au chapitre 1.

⁴⁴⁴ Voir par ex. M. Burri, « [Les politiques de diversité culturelle à l'ère de l'IA](#) », Cappello M. (éd.), *IRIS Spécial 2020-2, L'intelligence artificielle dans le secteur audiovisuel*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2020, p. 73 à 90.

⁴⁴⁵ Voir par ex. C. Gros et D. Gros, [Generative ai: A concise Primer for non-experts](#), Université de Bocconi, 2023.

⁴⁴⁶ Voir par ex. B. Jovanovic et P. L. Rousseau, « [General Purpose technologies](#) », P. Aghion et S. N. Durlauf (éd.), *Handbook of Economic Growth*, Elsevier, 2005, p. 1182–1224 ; J. Manyika et al., [Disruptive technologies: Advances That Will Transform Life, Business, and the Global Economy](#), McKinsey Global Institute, 2013.

⁴⁴⁷ Au niveau historique, l'imprimerie est également un exemple de technologie à usage général. Elle a tout d'abord été utilisée comme un moyen de rendre la Bible accessible, mais par la suite elle est devenue un outil grâce auquel les instigateurs de la Réforme ont pu imprimer leurs pamphlets pour diffuser leurs idées et élargir le mouvement. L'imprimerie a également contribué à déclencher la révolution scientifique et les Lumières en faisant largement connaître les recherches et les découvertes. Parmi les effets indirects, on peut noter la croissance accélérée des villes. Certains historiens attribuent l'essor rapide de l'Europe et de son influence à l'échelle mondiale, et le déclin des pays musulmans après le XV^e siècle au fait que l'imprimerie a été adoptée rapidement en Europe et beaucoup plus lentement dans les économies musulmanes. Voir J. Manyika et al. (2013), *ibid.*, p. 25 ; J. E. Dittmar, « [Information Technology and Economic change: The impact of the Printing Press](#) », *The Quarterly Journal of Economics* 126 (2011), 1133–1172.

⁴⁴⁸ Sur les technologies à usage général, voir par ex. E. Brynjolfsson, D. Li et L. Raymond, [Generative AI at Work](#), NBER WorkingPaper 31161 (2023). Sur l'IA en tant que technologie à usage général, voir par exemple le préambule du règlement de l'UE sur l'IA.

⁴⁴⁹ Voir par ex. Y. Benkler, « Growth-Oriented Law for the Networked information Economy: Emphasizing Freedom to Operate over Power to appropriate », Kauffman Taskforce on Law, Innovation and Growth (éd.), *Rules for Growth: Promoting Innovation and Growth through Legal Reform*, Kauffman Foundation (2011), 313–342 ; N. Helberger, J. Pierson et T. Poell, « [Governing Online Platforms: From Contested to Cooperative](#)



8.2. Impact de l'IA sur la liberté d'expression, le pluralisme des médias et la diversité culturelle

8.2.1. Remarques préliminaires : en quoi consiste le changement ?

Il est important de comprendre que l'IA conventionnelle et l'IA générative ont toutes deux des répercussions majeures sur la disponibilité, la diffusion et la consommation des contenus, mais que ces répercussions peuvent différer et se recouper mutuellement. En ce qui concerne les systèmes de recommandation basés sur l'IA, l'élément déterminant est l'altération de la dynamique de l'espace médiatique et l'émergence de ces systèmes en tant qu'intermédiaires critiques avec des fonctions éditoriales distinctes. Cela contraste fortement avec l'environnement hors ligne/analogique, où les fonctions éditoriales sont regroupées au sein d'un seul organisme (qu'il s'agisse d'un organe de presse ou de radiodiffusion), qui est également investi d'un certain mandat en vertu de la réglementation – notamment celui de présenter des contenus locaux et nationaux et de garantir la qualité et la fiabilité des informations. À l'instar de pratiquement toutes les politiques nationales en matière de médias et de culture, ce modèle traditionnel repose également sur le postulat selon lequel la diversité de l'offre se reflète dans la diversité de la consommation – tout en présumant que cela a des effets bénéfiques sur la formation de l'opinion, la participation politique et l'engagement culturel.

Les plateformes algorithmiques basées sur l'IA, telles que notamment les sites de réseaux sociaux, ont complètement modifié le paysage et provoqué des « changements radicaux dans la composition et la consommation des produits médiatiques »⁴⁵⁰. Ces « nouveaux éditeurs » sont multiples, dispersés et distribués⁴⁵¹ et en définitive, ce sont leurs algorithmes embarqués⁴⁵² qui définissent le nouvel espace médiatique et conditionnent l'exercice de la liberté d'expression, tant dans sa dimension active que passive⁴⁵³. Ce qui a par ailleurs des répercussions plus profondes sur le pluralisme et la diversité.

[Responsibility](#) », *The Information Society*, 2017 ; B. Judge, M. Nitzberg, S. Russell, « [When Code Isn't Law: Rethinking Regulation for Artificial Intelligence](#) », *Policy and Society* (2024), 43 [advance article].

⁴⁵⁰ R. Kleis Nielsen, R. Gorwa et M. de Cock Buning, [What Can Be Done? Digital Media Policy Options for Strengthening European Democracy](#), Reuters Institute Report, 2019.

⁴⁵¹ M. Latzer, K. Hollnbuchner, N. Just et F. Saurwein, « [The Economics of Algorithmic Selection on the Internet](#) », J. M. Bauer et M. Latzer (éd.), *Handbook on the Economics of the Internet*, Edward Elgar (2016), 395–425.

⁴⁵² Pour une définition complète des algorithmes, voir M. Latzer et N. Just, [Governance by and of Algorithms on the Internet: Impact and Consequences](#), Oxford Research Encyclopedia, Communication, Oxford University Press, 2020.

⁴⁵³ J. Balkin, [Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation](#), *UC Davies Law Review* 51 (2018), 1149–1210.



8.2.2. Impact sur la diffusion et la consommation des contenus

Dans ce contexte, il convient tout d'abord d'identifier les interférences possibles avec l'autonomie individuelle et la liberté de choix des utilisateurs. Alors que les systèmes de sélection et de recommandation basés sur l'IA réduisent les coûts de recherche et d'information et peuvent faciliter l'orientation sociale⁴⁵⁴, ils risquent également de favoriser l'émergence de risques sociaux, tels que des « menaces pour les droits et libertés fondamentaux, ainsi qu'un impact sur la médiation des réalités et le développement ultérieur des gens⁴⁵⁵. » Dans ce contexte, la personnalisation croissante du filtrage des médias est également importante, car elle est basée sur (et biaisée par) les profils des utilisateurs sur les réseaux sociaux et les expériences antérieures (par ex. avec les choix de « J'aime » et « Amis » ou l'historique de navigation, y compris la localisation), et en définitive, elle « promeut des contenus de proximité ainsi que des contenus socialement et idéologiquement familiers⁴⁵⁶. » Cela retient les utilisateurs au sein d'un cercle familial, en nourrissant leur curiosité avec des contenus de même type, ce qui renforce les opinions existantes et évacue les avis contradictoires⁴⁵⁷. On peut bien entendu affirmer que tel était le cas avec les médias traditionnels, puisque les gens étaient spontanément attirés par les contenus qu'ils avaient appréciés auparavant - néanmoins, la différence essentielle dans l'environnement actuel, c'est que les utilisateurs ne voient *plus que* ce type de contenus et que leur capacité d'exercer activement un choix est extrêmement limitée ou manipulée. C. P. Hoffman *et al.* expliquent que les réseaux sociaux ne font qu'exacerber cet effet en cultivant deux aspects de l'« homéophilie », à savoir la similitude des pairs et des contenus⁴⁵⁸.

À cet égard, il convient de garder à l'esprit qu'en dépit d'une légère baisse de l'utilisation des sites de réseaux sociaux comme point d'entrée de contenus et des variations entre les différents pays⁴⁵⁹, ils jouent toujours un rôle majeur en tant que contrôleurs d'accès. Cela renforce le poids de l'homéophilie et des représentations potentiellement tendancieuses tout en illustrant clairement le pouvoir de quelques opérateurs et l'impact considérable de leurs décisions – décisions qui sont notamment prises dans une optique essentiellement commerciale⁴⁶⁰, hors de tout mandat spécifique visant à garantir certains droits et/ou valeurs.

⁴⁵⁴ M. Latzer *et al.* (2020).

⁴⁵⁵ *Ibid.*, p. 29-30.

⁴⁵⁶ C. P. Hoffman, C. Lutz, M. Meckel et G. Ranzini, *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), 1360–1381.

⁴⁵⁷ *Ibid.* Parmi les ouvrages plus anciens, voir par ex. C. R. Sunstein, *Going to Extremes: How Like Minds Unite and Divide*, Oxford University Press, 2009 ; E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, Viking, 2011.

⁴⁵⁸ *Ibid.* ; voir aussi J. A. Tucker *et al.*, *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*, Hewlett Foundation, 2018.

⁴⁵⁹ Pour une analyse par pays, voir Reuters Institute, *Digital News Report 2024*, Oxford, 2024.

⁴⁶⁰ À titre d'exemple, Meta a essayé de réduire la place des informations sur Facebook, Instagram et Threads et a restreint la promotion algorithmique des contenus politiques. Par ailleurs, Meta a également réduit le



Il convient également de souligner l'explosion des « fausses informations » diffusées à des fins commerciales ou politiques, qui ont de graves répercussions sur les démocraties⁴⁶¹, mais aussi sur d'autres valeurs fondamentales, comme la santé publique⁴⁶². À mesure que la désinformation se généralise sur les plateformes, l'amplification algorithmique peut également renforcer son impact négatif, « en légitimant les récits fallacieux, conspirationnistes et mensongers. En ce sens, la recommandation algorithmique de contenus fallacieux peut entraver l'apprentissage mutuel, créer une illusion de soutien social, et reproduire ainsi des biais (dés)informationnels⁴⁶³. » Les nouvelles opportunités liées à l'IA ne font que permettre des versions plus sophistiquées de la falsification de la réalité, en particulier dans les images et les vidéos ; il existe également des risques liés aux erreurs et aux hallucinations auxquelles les systèmes d'IA sont souvent exposés⁴⁶⁴.

Dans ce contexte marqué à la fois par la désinformation, les fausses informations et les biais algorithmiques, il convient de noter que malgré un léger tournant vers les modèles d'abonnement payant aux sites d'information⁴⁶⁵, la grande majorité de la consommation en ligne se fait encore sur un petit nombre de plateformes soutenues en grande partie par la publicité, ce qui les contraint à lutter en permanence pour capter l'attention du public⁴⁶⁶. Alors que certains contenus agrégés proviennent de médias traditionnels⁴⁶⁷, ce qui peut infirmer certaines critiques récurrentes reprochant aux agrégateurs d'amplifier l'impact de sources non traditionnelles peu fiables⁴⁶⁸, il convient de noter que ce n'est pas le volume des contenus qui augmente, mais simplement leurs canaux de diffusion. Il est également pertinent de s'interroger sur la diversité effective des contenus disponibles, puisque certains d'entre eux ne sont que de simples divertissements réutilisés ou des produits à caractère commercial (comme les *vlogs* d'influenceurs) n'ayant que peu de valeur en termes d'enrichissement de la vie culturelle

soutien au secteur de l'information, notamment en ne renouvelant pas les contrats et en supprimant son onglet Actualités dans un certain nombre de pays. Voir Reuters Institute (2024).

⁴⁶¹ Voir par ex. L. W. Bennett et S. Livingston, « [The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions](#) », *European Journal of Communication* 33 (2018), p. 122–139. V. Dan et al., « [Visual mis- and Disinformation](#) », *Journalism and Mass communication Quarterly* 98 (2021), p. 641-664.

⁴⁶² Voir par ex. M. Burri, « [Fake News in Times of Pandemic and Beyond: Exploring of the rationales for Regulating information Platforms](#) », K. Mathis et A. Tor (éd.), *Law and Economics of the Coronavirus Crisis*, Springer, 2022, p. 31–58.

⁴⁶³ D. Shin et al., « [Countering Algorithmic Bias and Disinformation and effective Harnessing the Power of AI in Media](#) », *Journalism and Mass communication Quarterly* 99 (2022), 887–907, p. 890.

⁴⁶⁴ Reuters Institute, 2024.

⁴⁶⁵ Sur l'ensemble des marchés, seulement un cinquième environ des personnes interrogées (22 %) déclarent utiliser les sites internet ou les applications d'information comme leur principale source d'information en ligne (10 % de moins qu'en 2018). Les éditeurs de quelques marchés d'Europe du Nord ont réussi à contrer cette tendance, mais partout dans le monde, les jeunes affichent moins d'intérêt pour les organes d'information que par le passé. Voir Reuters Institute, 2024.

⁴⁶⁶ Reuters Institute, 2024.

⁴⁶⁷ *Ibid.*

⁴⁶⁸ Alors que les médias traditionnels sont importants pour l'information (en particulier sur les réseaux sociaux tels que Facebook et X), Reuters Institute (2024) constate un intérêt croissant pour les commentateurs partisans, les influenceurs et les jeunes créateurs d'information, en particulier sur YouTube et TikTok. Cela est également lié à la tendance selon laquelle la vidéo devient une source plus importante d'informations en ligne, en particulier auprès des plus jeunes.



ou politique. Néanmoins, il convient de noter que les médias traditionnels ont réagi au processus d'agrégation technologique et offrent beaucoup plus de contenus en ligne que dans leurs versions imprimées ou radiodiffusées. En ce qui concerne plus particulièrement les informations, Reuters Institute constate que les organismes de presse traditionnels investissent massivement dans les réseaux sociaux et enregistrent, selon eux, des niveaux élevés de trafic, d'audience hors site et de nouveaux abonnés numériques⁴⁶⁹.

En dépit du fait que cela donne accès à une diversité de contenus sur un plus grand nombre de plateformes, ce qui peut également attirer les jeunes, il convient de souligner deux inconvénients possibles : le premier est lié au « risque de plateforme » qui en découle, car les entreprises traditionnelles sont fortement tributaires de quelques rares plateformes (qui, parallèlement aux opérateurs traditionnels Google et Meta, incluent désormais TikTok⁴⁷⁰). Le second concerne le fait que l'approche des organes de presse traditionnels privés à l'égard des réseaux sociaux est fortement marquée par des modèles économiques dépendants des circuits en place et axés sur la publicité, les abonnements, ou une combinaison des deux⁴⁷¹.

8.2.3. Impact sur la création de contenus

L'IA offre également des outils de mesure très sophistiqués en ce qui concerne la diffusion et la consommation de contenus, ce qui peut conduire à un biais privilégiant les reportages grand public (en particulier avec la récente progression des formats vidéo courts⁴⁷²) plutôt que les reportages critiques et plus complexes – alternative que certains auteurs résumant ainsi : « objectifs liés au nombre de clics contre objectifs éditoriaux⁴⁷³. » Par conséquent, les algorithmes pèsent sur les décisions dans les médias en prédisant les habitudes de consommation et les préférences du public⁴⁷⁴. Si dans certains domaines, cela peut être considéré comme bénéfique pour donner aux différents publics ce qu'ils veulent, dans d'autres domaines, comme pour les informations, cela peut s'avérer très problématique, car les informations locales et les affaires courantes sont ajustées en fonction des variables démographiques, sociales et politiques de communautés

⁴⁶⁹ Ils identifient trois objectifs stratégiques principaux qui déterminent les différentes façons dont les organismes de presse abordent les médias sociaux : 1) augmenter le trafic en ligne grâce aux renvois ; 2) accroître l'audience hors ligne grâce aux formats natifs et aux contenus diffusés ; 3) promouvoir les ventes d'abonnements numériques, souvent grâce aux contenus publicitaires sur Facebook.

⁴⁷⁰ Reuters Institute, 2024.

⁴⁷¹ A. Cornia, A. Sehl, D. A. Levy et R. K. Nielsen, *Private Sector News, social Media distribution, and Algorithm change*, Reuters Institute, 2018.

⁴⁷² Reuters Institute, 2024.

⁴⁷³ Voir par ex. T. Dodds *et al.*, « *Popularity-Driven Metrics: Audience Analytics and Shifting opinion Power to Digital Platforms* », *Journalism Studies* 23 (2023), 403–421 ; voir aussi N. Helberger, « *FutureNewsCorp, or How the AI Act Changed the Future of News* », *Computer Law and Security Review* 52 (2024), 105915.

⁴⁷⁴ P. M. Napoli, « *On Automation in Media Industries: Integrating Algorithmic Media production into Media Industries Scholarship* », *Media Industries Journal* 1 (2014), 33–38 ; voir également F. Saurwein, N. Just et M. Latzer, « *Governance of Algorithms: Options and Limitations* », *info* 17 (2015), 35–49.



spécifiques⁴⁷⁵. Globalement, cela ne permet pas de développer une offre durable de contenus locaux, régionaux et nationaux diversifiés⁴⁷⁶, qui risquent ainsi de devenir très difficiles à trouver parce qu'ils sont ou seront marginalisés sur les plateformes en ligne⁴⁷⁷.

Comme nous l'avons déjà mentionné, la production de contenus IA est un phénomène relativement nouveau. Il y a bien sûr quelques « malheureux » exemples antérieurs avec la première génération de *content farms* (fermes de contenus), qui, sur la base de données des moteurs de recherche (tels que les termes de recherche les plus populaires, les ventes publicitaires et les contenus disponibles), produisaient des contenus rapidement et à moindre coût afin de répondre à cette demande⁴⁷⁸. Ce type de création de contenus est évidemment une marchandisation de l'information qui risque de porter préjudice à toutes les missions d'intérêt public des médias possibles et imaginables, notamment dans les domaines culturel et politique. Avec les nouvelles opportunités liées aux systèmes d'IA, la transformation au niveau de la production de contenus est d'une portée beaucoup plus large. À titre d'exemple, un livre blanc de l'Université d'Amsterdam montre en quoi l'IA transforme le journalisme et les médias – ce qui va des nouvelles façons de soutenir les investigations des journalistes jusqu'à l'aide à l'écriture et aux nouvelles formes de contact et d'interaction avec le public, notamment avec la diffusion personnalisée de contenus ou l'utilisation d'agents virtuels⁴⁷⁹. Pour ce qui est de l'aspect positif, cela renforce l'efficacité de la production de contenus (par exemple au regard des activités fastidieuses, telles que la transcription, la relecture et la mise en page) et permet également de lutter contre certains effets négatifs susmentionnés, notamment par la vérification des faits et la détection des hypertrucages⁴⁸⁰. À terme, cela devrait permettre de transformer radicalement la production et la diffusion de contenus d'une façon qui est difficilement imaginable à l'heure actuelle⁴⁸¹.

Mais il y a aussi le revers de la médaille, qui comporte un certain nombre d'inconvénients. L'un des principaux inconvénients est la dépendance accrue à l'égard des solutions basées sur l'IA et l'influence connexe des entreprises technologiques sur la production de contenus, ce qui touche de façon déterminante tous les aspects de la

⁴⁷⁵ P. M. Napoli, *ibid.*, p. 34. Reuters Institute (2024) constate que les éditeurs se focalisent peut-être trop sur la mise à jour des informations et ne consacrent pas assez de temps à présenter différentes perspectives sur certains problèmes ou à relayer des événements pouvant donner lieu ponctuellement à un certain optimisme. En ce qui concerne les sujets traités, il constate que le public est satisfait des informations politiques et sportives qu'il reçoit, mais fait état de lacunes au niveau des actualités locales, ainsi que des informations relatives à la santé et à l'éducation.

⁴⁷⁶ Voir p. ex. P. M. Napoli, M. Weber, K. McCollough et Q. Wang, *Assessing Local Journalism: News Deserts, Journalism Divide, and the Determinants of the Robustness of Local News*, News Measures Research Project, août 2018.

⁴⁷⁷ M. Burri, *Discoverability of Local, Regional and National Content*, A Thought Leadership Paper for the Canadian Commission for UNESCO and the Canadian Heritage, février 2019.

⁴⁷⁸ P. M. Napoli, 2014, p. 35.

⁴⁷⁹ A. Schjøtt Hansen *et al.*, « *Initial White Paper on the Social, Economic, and Political Impact of Media AI Technologies* », *AI4Media*, 2023 ; voir aussi C. Beckett et M. Yaseen, « *Generating Change. A Global Survey of What News Organisations Are Doing with AI* », *JournalismAI*, 2023.

⁴⁸⁰ Schjøtt Hansen *et al.*, 2023, *ibid.* ; voir aussi N. Newman, *Journalism, Media, and Technology Trends and Predictions 2023*, Reuters Institute, 2024.

⁴⁸¹ Pour une réflexion édifiante sur les scénarios d'avenir, voir N. Helberger, 2024.



société. C'est un fait établi que, pour leur part, les entreprises technologiques poursuivent leurs propres objectifs, tels que l'efficacité, l'évolutivité et la rapidité⁴⁸². Mais elles s'inspirent également des valeurs spécifiques liées à leur juridiction (qui, à l'heure actuelle, est aux États-Unis) qui reflètent une conception solidement ancrée du libéralisme économique, de la liberté d'expression ou de la vie privée et qui sont intégrées au sein même de la technologie⁴⁸³. En ce qui concerne la dépendance à l'IA et le pouvoir des entreprises technologiques, il peut y avoir un décalage avec les paradigmes des médias de service public et le rôle du journalisme dans une société démocratique qui prône le pluralisme et la diversité⁴⁸⁴. En ce qui concerne la place prépondérante des États-Unis dans le domaine de l'IA, il peut y avoir un décalage entre les valeurs sociétales communes à différents pays, qui est quelque peu implicite dans la technologie sous-jacente mais qui pourrait avoir d'importantes répercussions⁴⁸⁵.

8.2.4. Autres aspects à prendre en considération

On peut par ailleurs identifier un certain nombre d'autres aspects problématiques. Certains sont liés à l'opacité intrinsèque et au manque de sensibilisation contextuelle des applications de l'IA. Comme les systèmes d'IA ne peuvent pas comprendre le contexte sociétal, leur filtrage est souvent inexact. De ce fait, les faux positifs peuvent conduire à des restrictions injustifiées de la liberté d'expression, tandis que les faux négatifs peuvent avoir un effet dissuasif, conduisant à l'autocensure⁴⁸⁶. Par ailleurs, force est de constater que l'IA peut faciliter la surveillance et la censure, ce qui a de graves répercussions sur le droit de chercher et de recevoir des informations, ainsi que sur le pluralisme des médias⁴⁸⁷. Les inégalités en termes d'accès aux systèmes d'IA et d'utilisation de ces derniers posent de nouveaux défis – cela concerne les inégalités entre les régions et les pays, mais aussi entre les secteurs et les entreprises de ces secteurs. Dans le premier cas,

⁴⁸² *Ibid.*

⁴⁸³ Voir par ex. A. Chander, « [How Law Made Silicon Valley](#) », *Emory Law Journal* 63 (2014), p. 630 à 694 ; K. Klönick, « [The New Governors: The People, Rules, and Processes Governance Online Speech](#) », *Harvard Law Review* 131 (2018), 1598–1670.

⁴⁸⁴ F. M. Simon, « [Uneasy Bedfellows: AI in the News, Platform Companies and the issue of Journalist Autonomy](#) », *Digital Journalism* 10 (2022), 1832–1854 ; M. Moore et D. Tambini (éd.), *Digital Dominance: The Power of Google, Amazon, Facebook et Apple*, Oxford University Press, 2018 ; R. Kleis Nielsen et S. A. Ganter (éd.), *The Power of Platforms Shaping Media and Society*, Oxford University Press, 2022 ; J. van Dijck, T. Poell et Ma. de Waal, *The Platform Society: Public Values in a Connective World*, Oxford University Press, 2018.

⁴⁸⁵ L'une des leçons tirées à cet égard concerne le domaine de la protection de la vie privée et des données à caractère personnel, car l'UE et les États-Unis ont des conceptions très différentes en la matière et ont donc adopté des cadres réglementaires très différents.

⁴⁸⁶ Julia Haas, *Liberté de la presse et intelligence artificielle* (Bureau du Représentant pour la liberté des médias de l'OSCE, 2020) ; voir aussi N. Helberger *et al.*, *Artificial Intelligence – Intelligent Politics: Challenges and Opportunities for Media and Democracy*, document de synthèse, Conférence ministérielle du Conseil de l'Europe, Chypre, 28–29 mai 2020.

⁴⁸⁷ Julia Haas, *ibid.*, se référant également au Haut-Commissaire des Nations Unies aux droits de l'homme, [A/HRC/39/29](#), 2018 ; S. Feldstein, « [The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression](#) », *Journal of Democracy* 30 (2019), 40–52. A. Ünver, « [Artificial Intelligence, Authoritarianism and the Future of Political Systems](#) », *Cyber Governance and Digital Democracy* 9, 2018.



il s'agit d'une amplification de la fracture numérique « originelle » entre pays développés et pays en voie de développement, avec de graves conséquences également pour la représentation linguistique, car les données d'apprentissage sont le plus souvent en anglais ou dans d'autres langues dominantes au sein d'un pays, ce qui ne manquerait pas de porter préjudice à la pratique et la maîtrise des langues minoritaires⁴⁸⁸. Ces effets discriminatoires peuvent également se retrouver dans des secteurs distincts, et les industries culturelles peuvent être particulièrement vulnérables à cet égard, car peu d'artistes et d'entrepreneurs savent actuellement utiliser les outils d'IA et, s'ils le font, ils deviennent tributaire d'une poignée de systèmes commercialisés⁴⁸⁹. Enfin, il existe également des inégalités au sein même du secteur des médias, où les médias locaux et régionaux dotés de budgets plus modestes sont à la traîne et ont du mal à tirer profit des opportunités liées à l'IA⁴⁹⁰. Cela présente une fois de plus certains risques en termes d'appauvrissement de la diversité, de fragmentation et de distorsion potentielle du débat public et culturel.

Globalement, on peut affirmer que nous sommes dans une phase de profonde mutation de l'espace médiatique qui comporte des défis spécifiques pour le pluralisme et la diversité. Mais on peut également tirer certains avantages de l'IA. L'accessibilité de l'information est un avantage majeur, de même que le fait de fournir des informations plus pertinentes et potentiellement plus conformes aux intérêts d'un public hétérogène⁴⁹¹. Cela concerne non seulement les contenus actuels, mais aussi ceux qui proviennent des archives, notamment du patrimoine culturel et des organismes de conservation de la mémoire⁴⁹². Comme indiqué précédemment, les outils d'IA permettent également de renforcer l'efficacité des processus journalistiques et éditoriaux, ce qui peut libérer des ressources pour un journalisme de qualité et des enquêtes critiques. En ce qui concerne ce dernier élément, l'IA permet également de vérifier les faits et de tester la fiabilité des sources et des informations. Les outils d'IA peuvent également aider les entreprises de médias à offrir de nouveaux services, notamment des contenus personnalisés et des conseils rédactionnels, ce qui peut également s'inscrire dans de nouveaux modèles financiers qui peuvent aider les petits médias⁴⁹³. Il convient également de noter qu'il reste beaucoup de chemin à parcourir avant que les systèmes d'IA ne deviennent pleinement semblables aux humains⁴⁹⁴. Les organismes de médias peuvent investir dans ce « délai » en rétablissant la confiance et en montrant que leur pratique du journalisme repose sur l'exactitude, l'impartialité et la transparence, et que les humains gardent la maîtrise du processus d'information. Cette démarche devrait être perçue de façon positive par les

⁴⁸⁸ Schjøtt Hansen *et al.*, 2023, p. 90.

⁴⁸⁹ Voir par ex. O. Kulesz, *Culture, machines et plateformes : L'intelligence artificielle et son impact sur la diversité des expressions culturelles*, rapport pour l'UNESCO, DCE/18/12.IGC/INF.4 (2018) ; B. Caramiaux, *The Use of Artificial Intelligence in the Cultural and Creative Sectors*, rapport PE 629.220 du Parlement européen, 2020.

⁴⁹⁰ Schjøtt Hansen *et al.*, 2023, p. 90.

⁴⁹¹ *Ibid.*

⁴⁹² B. Caramiaux (2020). Cela va néanmoins de pair avec des défis spécifiques. Voir B. Caramiaux, « *AI with Museums and Cultural Heritage* », *AI in Museums*, De Gruyter (2023), 117–130 ; Magdalena Pasikowska-Schnass et Young-Shin Lim, *Artificial Intelligence in the Context of Cultural Heritage and Museums: Complex Challenges and New Opportunities*, rapport PE 747.120 du Parlement européen, 2023.

⁴⁹³ Schjøtt Hansen *et al.*, 2023 ; Shin *et al.*, 2022.

⁴⁹⁴ Shin *et al.*, 2022, p. 902.



différents publics, puisqu'ils recherchent des contenus diversifiés et fiables dans un espace informationnel très dense et chaotique⁴⁹⁵.

8.3. Gérer les répercussions réelles et potentielles des systèmes d'IA sur le pluralisme et la diversité

Depuis trois ou quatre ans, on assiste à une vague d'initiatives visant à réguler les plateformes au regard du rôle crucial qu'elles jouent dans la société, notamment en ce qui concerne la médiation des opinions, la diffusion de fausses informations et la désinformation⁴⁹⁶. On assiste également à une nouvelle vague supplémentaire de réglementation qui cible spécifiquement les systèmes d'IA en vue de prévenir et limiter leurs risques⁴⁹⁷. Même si ces initiatives ne répondent pas (encore) spécifiquement aux préoccupations liées au pluralisme et à la diversité que nous venons d'évoquer dans le présent chapitre, les implications des nouvelles réglementations – par exemple en matière de transparence, de responsabilité et de normalisation des procédures – peuvent être décisives en ayant un impact concret sur le terrain⁴⁹⁸. Il faut espérer que les nouvelles réglementations ne se contenteront pas d'encadrer simplement l'utilisation de l'IA, mais soutiendront le développement de pratiques responsables en la matière⁴⁹⁹. Par ailleurs, la communauté internationale est nettement plus consciente de la nécessité de prendre des mesures, comme en atteste la récente Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit⁵⁰⁰.

Dans une démarche visant à faire face au nouvel environnement tributaire de l'IA tout en veillant, d'une part, à préserver les valeurs fondamentales de la société et, d'autre part, à profiter des avantages de l'IA afin de garantir un environnement culturel pluraliste, sain et durable, la nécessité d'une action plus ciblée a fait l'objet d'un débat. À titre d'exemple, Schjøtt Hansen *et al.* proposent des orientations pertinentes en ce qui

⁴⁹⁵ Reuters Institute, 2024.

⁴⁹⁶ Voir par ex. M. Burri, 2022. L'Union européenne (UE) est à la pointe de ces initiatives avec une réglementation générique des plateformes inscrite dans la législation sur les services numériques (DSA), qui instaure des mesures de transparence de grande envergure en matière de modération des contenus et de publicité, ainsi que des obligations juridiquement contraignantes ayant force exécutoire, en particulier pour les très grandes plateformes en ligne, d'évaluer et de compenser les risques systémiques menaçant les droits fondamentaux ou liés à la manipulation délibérée de leurs services, notamment en ce qui concerne la lutte contre la désinformation en ligne. De plus, il existe un certain nombre de dispositifs plus spécifiques, tels que le règlement européen sur la liberté des médias ou le Code de bonnes pratiques contre la désinformation. Voir par exemple M. Burri, 2022, *ibid.*

⁴⁹⁷ Sur ce point, l'UE est de nouveau à l'avant-garde avec son règlement sur l'IA, mais d'autres États ont également adopté une législation sur l'IA, ne serait-ce que de façon disparate. Voir par ex. T. Giardini et J. Fritz, « [The Anatomy of AI Rules: A systematic comparative analysis of AI Rules across the Globe](#) », *Digital Policy Alert*, 2024.

⁴⁹⁸ Voir par ex. N. Helberger, 2024 ; N. Helberger et N. Diakopoulos, « [The European AI Act and How it Matters for Research into ai in Media and Journalism](#) », *Digital Journalism* 11 (2022), 1751–1760.

⁴⁹⁹ N. Helberger, 2024.

⁵⁰⁰ [CM\(2024\)52-final](#), 17 mai 2024.



concerne les mesures réglementaires ou autres qui pourraient être prises⁵⁰¹. Ils soulignent, entre autres, la nécessité de développer des infrastructures alternatives de données et de contenus en s'éloignant du modèle actuel de marchandisation pour pouvoir s'adapter aux spécificités – par exemple du secteur des médias, de certaines valeurs sociétales intrinsèques à la communauté, et des langues diverses et minoritaires. Cela pourrait s'articuler avec la nécessité de disposer d'ensembles de données spécifiques au domaine, *open source* et non commerciaux pour l'entraînement des systèmes d'IA, ainsi que de données qui ciblent spécifiquement les « valeurs européennes » intégrées. Par ailleurs, il semble nécessaire de se doter de bonnes pratiques en ce qui concerne l'obligation d'information, l'utilisation responsable des systèmes d'IA dans le secteur des médias, ainsi que la « diversité dès la conception »⁵⁰². Bien entendu, il ne faut pas oublier qu'à défaut de dispositifs de financement et d'incitation appropriés, il est peu probable qu'aucune de ces mesures ne se concrétise. Dans le même temps, lors de l'adoption d'une de ces mesures, il conviendra de trouver un certain équilibre pour ne pas entraver l'innovation – ce point a d'ailleurs retenu l'attention à l'occasion du Règlement sur l'IA de l'UE⁵⁰³.

8.4. Conclusion

Ce chapitre a présenté quelques aspects de la profonde mutation à l'œuvre dans l'espace médiatique sous l'effet de l'accélération des avancées technologiques et, plus récemment, de l'avènement et la diffusion à grande échelle de la GenAI. Cette évolution est radicale et a des répercussions majeures sur la création, la diffusion, l'utilisation, la réutilisation et la consommation des contenus, répercussions dont nous n'avons pas encore pleinement pris la mesure. Cela a également un impact sur le pluralisme et la diversité, qui sont des piliers fondamentaux de nos sociétés démocratiques. Pour sauvegarder nos droits fondamentaux et nos valeurs sociétales, il est nécessaire d'agir, mais la voie à suivre à cet égard n'est ni évidente, ni facile, car l'environnement étant très fluide, on ne peut exclure des effets indésirables. La prise de conscience est le premier pas dans la bonne direction – non seulement en termes de compréhension des possibilités technologiques, mais aussi des valeurs fondamentales de notre société qu'il nous faut préserver.

⁵⁰¹ Schjøtt Hansen *et al.*, 2023.

⁵⁰² *Ibid.* Sur le thème de la « diversité dès la conception », voir N. Helberger, « [Diversity Label: Exploring the Potential and Limits of a Transparency Approach to Media Diversity](#) », *Journal of Information Policy* 1 (2011), p. 337 à 369 ; N. Helberger, « [Diversity by Design](#) », *Journal of Information Policy* 1 (2011), p. 441-469.

⁵⁰³ Certains rapports soutiennent, par exemple, qu'il n'est pas nécessaire d'intervenir et que le Règlement sur l'IA de l'UE risque d'étouffer l'innovation en matière de GenAI en Europe. Voir par ex. P. R. Abecasis *et al.*, *Generative Artificial Intelligence: The Competitive Landscape*, Copenhagen Economics, 2024.

PART IV – L'IA générative dans nos vies : l'importance de l'éducation et de la sensibilisation

L'IA générative fait désormais partie de nos vies, et nous devons apprendre à tirer le meilleur parti de son potentiel en comprenant comment l'utiliser efficacement. Si l'on met de côté les défis réglementaires qu'elle pose et que l'on se tourne vers l'avenir, il est nécessaire d'éduquer les gens à l'IA et d'en améliorer la connaissance dans la vie de tous les jours. Cela soulève des questions sur l'évolution du rôle des institutions publiques.

L'école devrait-elle enseigner aux jeunes enfants comment utiliser et évaluer les contenus générés par l'IA et, dans l'affirmative, au moyen de quels outils ? La machine pourrait-elle à terme assister les enseignants en les aidant à personnaliser le contenu en fonction de leur public ? Pourrait-elle posséder une sensibilité comparable à celle de l'homme ?

Les médias de service public pourraient également jouer un rôle dans la sensibilisation à l'IA. Par exemple, un futur assistant vocal pourrait diffuser des informations quotidiennes de manière sélective, en utilisant potentiellement des informations qui ne sont pas vérifiées ou exactes. Les médias de service public et leurs services d'information devraient-ils collaborer avec les entreprises d'IA générative pour garantir l'exactitude des informations, une touche humaine dans le journalisme ainsi que le pluralisme des médias ? Quelles seraient les conséquences si l'information et le journalisme n'étaient pas créés ou diffusés par des êtres humains ?

Ces questions restent ouvertes, et, alors que nous évoluons dans ce paysage en pleine mutation, seul l'avenir nous apportera des réponses.



9. Le monde de demain : les textes juridiques sont-ils à l'épreuve de l'intelligence artificielle et des défis de l'audiovisuel ?

Mark Cole, directeur des affaires académiques, Institut du droit européen des médias et professeur de droit des médias et des télécommunications, Université du Luxembourg et Sandra Schmitz-Berndt, projet LAIWYERS, Université du Luxembourg

9.1. Récapitulatif : approches réglementaires existantes et à venir

Face au développement rapide des capacités de l'IA, une réglementation a vu le jour – assez rapidement dans certaines parties du monde – pour tenter de faire face aux risques et aux défis liés à l'utilisation de l'IA. Les premières initiatives se sont principalement concentrées sur une utilisation éthique de l'IA et ont cherché à introduire des principes auxquels les déployeurs devraient adhérer⁵⁰⁴. Ces principes portaient essentiellement sur la technologie sous-jacente et concernaient plus généralement l'équité, l'exactitude et la transparence.

⁵⁰⁴ Sur l'évolution de la réglementation de l'IA, voir Mark D. Cole, « [AI Regulation and Governance on a Global Scale : An Overview of International, Regional and National Instruments](#) » (2024) 1(1) *AIRe* 126.



9.1.1. En commençant par les recommandations : approches initiales de l'OCDE et de l'UNESCO

La Recommandation de l'OCDE sur l'intelligence artificielle de mai 2019⁵⁰⁵ peut être considérée comme le premier « point de référence mondial pour une IA digne de confiance⁵⁰⁶ ». La recommandation vise à favoriser l'innovation et la confiance dans l'IA en promouvant la gestion responsable d'une IA digne de confiance tout en garantissant le respect des droits de l'homme et des valeurs démocratiques. La recommandation contient notamment une définition de l'IA qui a déjà été révisée en novembre 2023 pour englober l'IA générative⁵⁰⁷. Complétant les normes existantes de l'OCDE dans des domaines tels que la protection de la vie privée, la gestion des risques liés à la sécurité numérique et la conduite responsable des entreprises, la recommandation introduit des principes fondamentaux fondés sur des valeurs pour assurer le déploiement, le développement et l'utilisation d'une IA digne de confiance, à savoir : (1) la croissance inclusive, le développement durable et le bien-être ; (2) les valeurs centrées sur l'humain et l'équité ; (3) la transparence et l'explicabilité ; (4) la robustesse, la sécurité et la sûreté ; et (5) la responsabilité pour le bon fonctionnement des systèmes d'IA et le respect des principes. Ces principes sont devenus un point de référence déterminant dans les cadres politiques ultérieurs, notamment les déclarations politiques telles que les principes du G7 et du G20 en matière d'IA⁵⁰⁸, ainsi que la déclaration de Bletchley⁵⁰⁹ publiée par les États participant au sommet sur la sécurité de l'IA tenu en novembre 2023 au Royaume-Uni.

La Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle⁵¹⁰, qui constitue le premier cadre normatif mondial, aborde les caractéristiques des systèmes d'IA qui revêtent une importance capitale sur le plan éthique. Les objectifs de gouvernance de la recommandation portent sur la manière dont les contenus générés par IA sont utilisés du point de vue, entre autres, de la durabilité, de l'égalité des genres ainsi que dans le contexte de l'emploi, afin d'assurer un développement sain de l'IA. Comme cela a déjà été évoqué dans les chapitres précédents de cette publication, l'IA soulève des préoccupations éthiques profondes et multiples⁵¹¹, dont les plus importantes concernent ses biais, l'équité, les menaces posées aux droits de l'homme, la répartition des services et

⁵⁰⁵ OCDE, [Recommandation du Conseil sur l'intelligence artificielle](#) (2019) OECD/LEGAL/0449.

⁵⁰⁶ Lucia Russo et Noah Oder, [How Countries Are Implementing the OECD Principles for Trustworthy AI](#) (oecd.ai, 31 octobre 2023).

⁵⁰⁷ Selon la recommandation révisée, un « système d'IA » est défini comme « un système automatisé qui, pour des objectifs explicites ou implicites, déduit, à partir d'entrées reçues, comment générer des résultats en sortie tels que des prévisions, des contenus, des recommandations ou des décisions qui peuvent influencer sur des environnements physiques ou virtuels. Différents systèmes d'IA présentent des degrés variables d'autonomie et d'adaptabilité après déploiement ».

⁵⁰⁸ G20, [Principes du G20 sur l'IA](#), Annexe à la déclaration ministérielle du G20 sur le commerce et l'économie numérique (juin 2019) ; G7, [Déclaration des dirigeants du G7 sur le processus d'Hiroshima en matière d'IA](#) (1^{er} décembre 2023).

⁵⁰⁹ [Déclaration de Bletchley](#) des pays participant au sommet sur la sécurité de l'IA, les 1^{er} et 2 novembre 2023 (1^{er} novembre 2023).

⁵¹⁰ UNESCO, [Recommandation sur l'éthique de l'intelligence artificielle](#) (2022) SHS/BIO/PI/2021/1.

⁵¹¹ Voir également, pour une réflexion initiale, Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter et Luciano Floridi, « [The Ethics of Algorithms : Mapping the Debate](#) » (2016) 3(2) *Big Data Soc.*



des biens et les mouvements économiques, risquant de renforcer les inégalités existantes. L'UNESCO a cherché à fournir une « boussole éthique » en répondant à ces défis par le biais de cette recommandation. Une fois encore, celle-ci est conçue comme un point de référence qui sera concrétisé par de futures obligations.

9.1.2. Vers un droit contraignant : les avancées au sein du Conseil de l'Europe et de l'UE

Parallèlement à ces instruments non contraignants de l'OCDE et de l'UNESCO, les premiers textes « de droit dur » dans ce domaine ont récemment vu le jour en Europe. Ceux-ci vont encore plus loin et se démarquent par la rapidité de leur adoption et par leur ambition en termes de portée territoriale et sur le fond. Il s'agit de la convention-cadre sur l'IA du Conseil de l'Europe et du Règlement sur l'IA de l'Union européenne.

9.1.2.1. Convention-cadre sur l'IA du Conseil de l'Europe

S'appuyant sur les instruments non contraignants et les déclarations politiques mentionnés ci-dessus, le Comité des ministres du Conseil de l'Europe a formellement adopté, le 17 mai 2024, le premier traité de droit international public sur l'IA, la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit⁵¹². Si le Conseil de l'Europe rassemble ses 46 États membres, d'autres États, dont les États-Unis et le Japon, ainsi que des représentants de la société civile, du monde universitaire et de l'industrie ont également été activement impliqués dans le processus de rédaction. Par conséquent, lorsque la convention a été signée le 5 septembre 2024, ce sont non seulement les États membres du Conseil de l'Europe⁵¹³, mais aussi Israël, l'UE et les États-Unis qui ont signé et choisi d'être juridiquement liés par la convention⁵¹⁴. La convention entrera en vigueur peu après le dépôt des cinq premiers instruments de ratification ou instruments comparables des signataires, dont trois doivent être des États membres du Conseil de l'Europe. On peut s'attendre à ce que de nombreux autres États viennent maintenant signer la convention. En outre, le Conseil de l'Europe considère le domaine de l'intelligence artificielle comme une « priorité transversale⁵¹⁵ » et le préambule de la convention précise qu'elle a « valeur de cadre et qu'elle peut être complétée par d'autres instruments destinés à traiter des questions spécifiques liées aux activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle », ce qui signifie que d'autres textes sectoriels pourraient être adoptés par le Conseil de l'Europe.

⁵¹² Conseil de l'Europe, [Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit](#), STCE n° 225.

⁵¹³ Andorre, Géorgie, Islande, Norvège, République de Moldova, Royaume-Uni et Saint-Marin.

⁵¹⁴ Voir [ici](#) la liste des signatures et des ratifications (pas encore de ratification à ce jour car le traité vient d'être ouvert à la signature).

⁵¹⁵ Conseil de l'Europe, [Le Conseil de l'Europe et l'intelligence artificielle](#) (mars 2023), p. 3.



En mettant l'accent sur l'engagement à protéger les droits de l'homme, la convention-cadre adopte une approche fondée sur les risques et cherche à compléter les cadres existants en matière de droits de l'homme afin de garantir que les niveaux de protection des droits de l'homme, de la démocratie et de l'État de droit des États parties soient également garantis au vu des défis actuels et futurs soulevés par l'IA⁵¹⁶. Ainsi, la convention-cadre formule des principes fondamentaux visant à combler les lacunes juridiques résultant des avancées technologiques. Elle demande aux parties de s'assurer que leur cadre existant en matière de droits de l'homme répond de manière suffisante et efficace aux risques et aux défis de l'IA et, si ce n'est pas le cas, de déterminer comment l'adapter afin de maintenir les niveaux de protection existants. En utilisant l'instrument juridique d'une convention-cadre plutôt qu'une réglementation complète sous la forme d'une convention créant des droits et des obligations, l'instrument énonce de grands principes, des valeurs fondamentales et des domaines d'action, en vue d'inscrire les normes existantes en matière de droits de l'homme, de démocratie et d'État de droit dans le contexte de l'IA. Ces grands principes permettent une interprétation pouvant s'adapter aux défis d'un monde en évolution, alors que, en particulier dans le domaine technologique, des dispositions juridiques détaillées peuvent souvent gêner la mise en œuvre de l'instrument lorsque le contexte change⁵¹⁷. Néanmoins, la convention-cadre ne se limite pas à donner des éléments substantiels à prendre en compte par les parties qui la ratifient, mais prévoit également la mise en place d'une supervision efficace et de structures institutionnelles à cet effet.

Les principes fondamentaux inscrits dans la convention-cadre s'écartent légèrement des principes de l'OCDE en ce qu'ils incluent également la dignité humaine et l'autonomie individuelle, la protection de la vie privée et des données, l'égalité et la non-discrimination, la fiabilité et la sécurité de l'innovation. La plupart de ces principes supplémentaires avaient déjà été inclus dans les valeurs et principes de la recommandation de l'UNESCO. Si la dignité humaine et les droits fondamentaux sont davantage au cœur des travaux du Conseil de l'Europe, l'inclusion parallèle de l'innovation sûre peut paraître plus surprenante dans le contexte d'un document émanant de cette organisation. Les principes sont complétés par l'obligation pour les États d'adopter un cadre de gestion des risques et des impacts et par des obligations relatives à la mise en œuvre de la convention. Il convient de noter que, bien que la convention exige des parties qu'elles mettent en place des mécanismes d'application et de contrôle, son efficacité future dépendra en grande partie de la mise en œuvre au niveau national d'exigences formulées de manière générale. Afin de renforcer la coopération internationale et de faciliter les efforts d'harmonisation de la gouvernance de l'IA au niveau mondial, la convention-cadre reprend la définition d'un système d'intelligence artificielle issue de la dernière révision adoptée par l'OCDE. La convention-cadre est non seulement ouverte à la ratification par des pays non membres du Conseil de l'Europe, mais ceux-ci avaient déjà été intégrés dans le processus de rédaction, compte tenu de la

⁵¹⁶ Voir le mot du Président à la suite de l'adoption du texte, [Mot du Président](#).

⁵¹⁷ Voir le [Rapport sur l'intelligence artificielle \(2019\) T-PD\(2018\)09Rev](#) du Comité consultatif du Conseil de l'Europe sur la [Convention pour la protection des personnes à l'égard du traitement des données à caractère personnel \(Convention 108\)](#).



localisation de nombreuses entreprises innovantes dans ce domaine, en particulier les États-Unis. Ainsi, la convention-cadre pourrait devenir la « norme mondiale de droit contraignant », avec une sorte d'« effet Strasbourg » comparable à l'« effet Bruxelles » de certains textes législatifs de l'Union européenne concernant la sphère numérique⁵¹⁸.

Bien que la convention-cadre couvre le secteur public et les secteurs privés agissant au nom des autorités publiques, elle laisse aux parties une grande marge de manœuvre dans l'application des dispositions au secteur privé. Cette approche « à la carte⁵¹⁹ » permet aux parties de choisir de ne pas appliquer les principes et les obligations de la convention-cadre directement aux activités du secteur privé⁵²⁰. En conséquence, la pertinence de la convention-cadre (qui s'applique de manière horizontale) dans le secteur des médias audiovisuels dépend en grande partie de la mise en œuvre nationale du Traité. S'il n'est pas obligatoire pour les États qui l'ont ratifiée d'en étendre l'application dans ce sens, la convention-cadre prévoit néanmoins l'obligation de prendre « des mesures appropriées » pour satisfaire à l'obligation de traiter les éventuels « risques et [...] impacts découlant des activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle par des acteurs privés⁵²¹ », ce qui signifie que les États ne sauraient ignorer l'importance de ce secteur lorsqu'ils appliquent les exigences découlant de la convention-cadre.

9.1.2.2. Le Règlement sur l'IA de l'Union européenne de 2024

Contrairement à la convention-cadre du Conseil de l'Europe qui repose sur des principes, le Règlement sur l'IA de l'UE⁵²² suit une approche axée sur le marché, qui établit des règles communes visant à harmoniser le marché intérieur de l'UE. Ce texte a été adopté sous la forme d'un règlement, ce qui en fait un instrument directement applicable et contraignant pour les 27 États membres de l'UE. Il est entré en vigueur le 1^{er} août 2024 et – pour sa plupart⁵²³ – deviendra applicable à partir du 2 août 2026 au terme d'une période de transition.

⁵¹⁸ Sur l'« effet Bruxelles », voir Mark D. Cole et Christina Etteldorf, « [The Implementation of the GDPR in Member States' Law and Issues of Coherence and Consistency](#) », in Inge Graef et Bart van der Sloot (éd.), *The Legal Consistency of Technology Regulation in Europe* (Hart Publishing, 2024), p. 131-156. Pour un instrument normatif antérieur du Conseil de l'Europe concernant l'environnement en ligne, voir également la [Convention sur la cybercriminalité](#) de 2001, STE n° 185, qui a été signée et ratifiée par 76 États, dont les États-Unis, le Japon, l'Australie, le Brésil et le Nigéria.

⁵¹⁹ Voir Christopher Lamont, [The Council of Europe's draft AI Treaty : Balancing National Security, Innovation and Human Rights ?](#) (18 mars 2024), Global Governance Institute.

⁵²⁰ Dans ce cas cependant, il était prévu dès le stade de la rédaction que les approches de ces parties évolueraient au fil du temps, à mesure que leurs stratégies de réglementation du secteur privé progresseraient. Voir le [Rapport explicatif](#), paragraphe 30.

⁵²¹ Article 3(1) de la convention-cadre.

⁵²² [Règlement \(UE\) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'intelligence artificielle et modifiant les règlements \(CE\) n° 300/2008, \(UE\) n° 167/2013, \(UE\) n° 168/2013, \(UE\) 2018/858, \(UE\) 2018/1139 et \(UE\) 2019/2144 et les directives 2014/90/UE, \(UE\) 2016/797 et \(UE\) 2020/1828 \(règlement sur l'intelligence artificielle\)](#), JO L 2024/1689, 12 juillet 2024.

⁵²³ Les règles concernant les pratiques interdites en matière d'IA seront par exemple déjà applicables le 2 août 2025.



Le Règlement sur l'IA, beaucoup plus détaillé, vise à favoriser le développement de systèmes d'IA sûrs et dignes de confiance dans l'ensemble du marché unique de l'UE par des acteurs privés et publics et à stimuler l'investissement et l'innovation en matière d'IA dans l'UE. Dans le même temps, la loi vise à garantir le respect des droits fondamentaux des citoyens de l'UE. À l'instar de la convention-cadre, le Règlement sur l'IA suit une approche fondée sur les risques selon laquelle les différents types de systèmes d'IA sont classés en fonction des risques qu'ils présentent et interdit certaines pratiques d'IA pour lesquelles ces risques sont considérés comme inacceptables. Au cœur du Règlement sur l'IA se trouvent les exigences en matière de sécurité des produits. À cet égard, le règlement fixe des exigences spécifiques pour les systèmes d'IA à haut risque et des obligations pour les exploitants de ces systèmes. Dans la mesure où le règlement s'applique à la fois au secteur privé et au secteur public, son impact est plus large. Il établit un degré de rigueur différent pour les normes applicables en fonction du niveau de risque. Par exemple, les obligations sont plus strictes lorsque des systèmes d'IA à haut risque sont mis en service par des autorités publiques. Certaines exigences sont prévues pour les systèmes d'IA à haut risque, notamment un système de gestion des risques, des règles de gouvernance des données, une documentation technique, la tenue de registres, des obligations de transparence et d'information, un contrôle humain, l'exactitude, la robustesse et la cybersécurité.

Ces exigences s'inscrivent clairement en lien avec l'acquis de l'UE dans le secteur numérique – avec lequel elles se recouperont après la mise en application du Règlement sur l'IA⁵²⁴. Les exigences en matière de transparence découlent notamment du RGPD⁵²⁵, du Règlement P2B⁵²⁶ ou du Règlement sur les services numériques (*Digital Services Act* – DSA)⁵²⁷. Les règles relatives à la cybersécurité ont pour leur part été mises en place, entre autres, avec le RGPD et les directives SRI⁵²⁸ et figurent aussi dans le futur règlement sur la cyberrésilience⁵²⁹. Les règles relatives aux données apparaissent quant à elle, au-delà du

⁵²⁴ Pour une vue d'ensemble, voir Christina Etteldorf, « Le cadre réglementaire de l'Union européenne », in : Cappello M. (éd.), *Transparence et responsabilité en matière d'algorithmes des services numériques*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2023, p. 14 et s.

⁵²⁵ [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 \(règlement général sur la protection des données\)](#), JO L 119/1, 4 mai 2016.

⁵²⁶ [Règlement \(UE\) 2019/1150 du Parlement européen et du Conseil du 20 juin 2019 promouvant l'équité et la transparence pour les entreprises utilisatrices de services d'intermédiation en ligne](#), JO L 186/57, 11 juillet 2019.

⁵²⁷ [Règlement \(UE\) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE \(règlement sur les services numériques\)](#), JO L 277/1, 27 octobre 2022.

⁵²⁸ [Directive \(UE\) 2016/1148 du Parlement européen et du Conseil du 6 juillet 2016 concernant des mesures destinées à assurer un niveau élevé commun de sécurité des réseaux et des systèmes d'information dans l'Union \(Directive SRI\)](#), JO L 194/1, 19 juillet 2016, qui va être remplacée par la [Directive \(UE\) 2022/2555 du Parlement européen et du Conseil du 14 décembre 2022 concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union, modifiant le règlement \(UE\) no 910/2014 et la directive \(UE\) 2018/1972, et abrogeant la directive \(UE\) 2016/1148 \(directive SRI 2\)](#), JO L 333/80, 27 décembre 2022.

⁵²⁹ Commission européenne, [Proposition de règlement du Parlement européen et du Conseil concernant des exigences horizontales en matière de cybersécurité pour les produits comportant des éléments numériques et](#)



RGPD, dans le Règlement sur les données⁵³⁰ et le Règlement sur la gouvernance des données⁵³¹. Lorsque ces actes législatifs traitent des algorithmes et des processus de prise de décision automatique, ils constituent des précédents également pour le Règlement sur l'IA. Cela inclut même la directive sur les Services de médias audiovisuels qui, depuis sa révision en 2018⁵³² et l'extension de son application aux plateformes de partage de vidéos (*video-sharing platforms* – VSP), fait également référence aux algorithmes⁵³³.

D'autres avancées en matière de politique et de réglementation de l'IA ont certainement été déclenchées par les nouveaux défis posés par les modèles d'IA à usage général (*general-purpose AI* – GPAI). Certains de ces défis ne sont devenus évidents qu'alors que le processus d'élaboration des textes mentionnés ci-dessus était déjà bien avancé. De fait, le sommet sur la sécurité de l'IA et la déclaration de Bletchley qui en a résulté sont une réponse directe aux capacités accrues des GPAI ; il en va de même pour le décret présidentiel d'octobre 2023 aux États-Unis⁵³⁴ qui préconise l'adoption de normes nationales en matière d'IA par le pouvoir exécutif afin de répondre aux capacités croissantes de l'IA et à leurs implications. Cependant, tant la convention-cadre que le Règlement sur l'IA traitent de l'IA générative. Point important dans le contexte des médias, la convention-cadre ainsi que le Règlement sur l'IA abordent la nécessité d'éviter le risque de manipulation et de tromperie dans le cadre de l'obligation de transparence relative à l'identification du contenu généré par les systèmes d'IA⁵³⁵.

[modifiant le règlement \(UE\) 2019/1020 COM/2022/454 final](#). Pour une vue d'ensemble, voir Mark D. Cole et Sandra Schmitz-Berndt, « [Towards an Efficient and Coherent Regulatory Framework on Cybersecurity in the EU : The Proposals for a NIS 2.0 Directive and a Cyber Resilience Act](#) », (2022) 1(1) *Applied Cybersecurity & Internet Governance*.

⁵³⁰ [Règlement \(UE\) 2023/2854 du Parlement européen et du Conseil du 13 décembre 2023 concernant des règles harmonisées portant sur l'équité de l'accès aux données et de l'utilisation des données et modifiant le Règlement \(UE\) 2017/2394 et la Directive \(UE\) 2020/1828 \(Règlement sur les données\)](#), JO L 2023/2854, 22 décembre 2023.

⁵³¹ [Règlement \(UE\) 2022/868 du Parlement européen et du Conseil du 30 mai 2022 portant sur la gouvernance européenne des données et modifiant le Règlement \(UE\) 2018/1724 \(Règlement sur la gouvernance des données\)](#), JO L 152/1, 3 juin 2022.

⁵³² [Directive \(UE\) 2018/1808 du Parlement européen et du Conseil du 14 novembre 2018 modifiant la Directive 2010/13/UE 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 303/69, 28 novembre 2018.

⁵³³ Notamment l'article 28 *ter* de la Directive SMAV pour les mesures que les États membres pourraient exiger des fournisseurs de VSP, voir à ce sujet Mark D. Cole et Christina Etteldorf, [Future Regulation of Cross-Border Audiovisual Content Dissemination](#), Schriftenreihe Medienforschung Bd. 84 (Nomos, 2024), p. 108 et s., p. 125 et s., p. 206 et s.

⁵³⁴ [Executive Order \(E.O.\) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence](#), 30 octobre 2023.

⁵³⁵ Article 6 de la convention-cadre ; voir également le rapport explicatif, paragraphes 59 et 104.



9.2. Les approches réglementaires à l'épreuve de la réalité : applicabilité et limites dans le cadre du secteur audiovisuel

Compte tenu de la diversité des initiatives présentées ci-dessus, on peut se demander si ces cadres sont adaptés aux caractéristiques du contexte de l'audiovisuel. Les chapitres 2 à 8 ont déjà mis en évidence un certain nombre de défis juridiques liés à l'utilisation de l'IA dans le secteur audiovisuel, tout en précisant qu'aucune règle sectorielle directement contraignante n'a été mise en place jusqu'à présent. Il est donc d'autant plus important de se demander dans quelle mesure la législation existante a un impact indirect sur les systèmes d'IA ou si les cadres nouvellement créés ont été conçus de manière à pouvoir répondre à ces défis et risques spécifiques du secteur⁵³⁶.

9.2.1. Aspects relatifs à la protection des données

Dans la mesure où il est fondé sur des données, l'apprentissage d'un système d'IA dépend inévitablement de vastes ensembles de données, ce qui peut entraîner des infractions à la législation sur la protection des données et de la vie privée. Comme l'explique en détail le chapitre 2, les supports de formation audiovisuels contiennent très souvent des données à caractère personnel, ce qui nécessite une base juridique – en vertu de la législation européenne – pour des activités telles que le moissonnage et la copie de contenu⁵³⁷.

Le respect des règles de protection des données va plus loin et concerne également les transferts de données de l'UE vers des pays tiers, pour lesquels des limitations sont en place afin de garantir que la protection des données à caractère personnel n'est pas compromise par le transfert de données vers des pays tiers qui ne disposent pas d'un niveau de protection adéquat par rapport à celui garanti dans l'UE, ou pour lesquels aucune autre mesure de protection n'est mise en place. Les ensembles de données utilisés pour entraîner, valider et tester les systèmes d'IA, l'interaction des individus avec les systèmes d'IA et le contenu généré par un système d'IA ont une dimension transfrontalière, notamment entre des personnes dans l'UE et des entreprises de traitement situées par exemple aux États-Unis ou dans des pays asiatiques. C'est pourquoi les limitations potentielles imposées par les règles interdisant les transferts de données sans garanties sont importantes. Les responsables s'accordent à dire, dans le cadre d'initiatives mondiales, qu'une collaboration plus poussée est nécessaire sur ces questions en particulier dans le contexte de l'IA générative⁵³⁸, notamment pour garantir

⁵³⁶ Voir également les contributions dans Cappello M. (éd.), *L'intelligence artificielle dans le secteur audiovisuel*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2020.

⁵³⁷ Voir le chapitre 2.3.1. Pour l'UE, voir l'article 6 du RGPD.

⁵³⁸ Voir la section 2.6 et la table ronde des autorités de protection des données et de la vie privée du G7, *Déclaration sur l'IA générative* (2023).



des niveaux de protection comparables susceptibles de faciliter les flux de données. Dans cette optique, par exemple, la table ronde des autorités de protection des données du G7⁵³⁹ étudie actuellement la meilleure façon de protéger la vie privée dans ce contexte, et exhorte les développeurs d'IA à veiller à ce que les informations personnelles créées par l'IA générative soient tenues exactes, complètes et à jour⁵⁴⁰, ce qui constitue une autre obligation en vertu de la loi de l'UE relative à la protection des données. Le considérant 9 du Règlement sur l'IA de l'UE précise que les règles harmonisées qu'il prévoit complètent le droit de l'UE existant, en particulier en ce qui concerne les droits fondamentaux et la protection des données. Parce qu'elle repose sur des droits fondamentaux, la convention-cadre souligne elle aussi l'importance de la protection des données à caractère personnel pour préserver le droit à la vie privée dans le monde numérique⁵⁴¹, à l'instar de la Recommandation de l'OCDE. La convention-cadre vient compléter les cadres existants en matière de droits de l'homme. En ce qui concerne la protection des données, il existe un instrument juridique contraignant au sein du Conseil de l'Europe avec la Convention modernisée pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel⁵⁴². La modernisation de la convention en 2019 a assuré un alignement avec la réforme de l'UE de 2016 et l'adoption du RGPD, et a également revu le régime des flux transfrontaliers de données à caractère personnel. Ce dernier vise à garantir que les données à caractère personnel sont protégées par des garanties appropriées, même si elles relèvent de la juridiction d'un pays qui n'est pas partie au texte⁵⁴³. La garantie d'une telle protection représente évidemment un défi si l'on considère que les États-Unis, où sont basées de nombreuses entreprises technologiques, ne sont pas partie à la Convention et ont une approche très différente en matière de limitation du traitement des données à caractère personnel. Cela souligne l'importance du travail du G7 dans ce domaine et de l'accord sur le transfert de données entre les États-Unis et l'UE. Pour des raisons évidentes, ceci est d'autant plus important pour l'industrie du contenu puisqu'une grande partie de la production audiovisuelle provient aujourd'hui des États-Unis.

9.2.2. Aspects relatifs aux droits de propriété intellectuelle

Comme le souligne plus en détail le chapitre 3, les données d'apprentissage comprennent souvent du matériel protégé par des droits d'auteur. Étant donné que les modèles d'IA générative nécessitent de grandes quantités de données, les techniques de fouille de

⁵³⁹ Cette table ronde réunit les autorités chargées du contrôle de la protection des données et de la vie privée des pays du G7 afin d'intensifier la coopération réglementaire et d'examiner les questions relatives à la sécurité et à la fiabilité du traitement transfrontalier des données à caractère personnel dans une société de plus en plus numérisée et mondialisée.

⁵⁴⁰ Voir *ibid.*

⁵⁴¹ Voir article 11 de la convention-cadre.

⁵⁴² Conseil de l'Europe, [Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel](#), STE 108+ (STE 223 qui, après son entrée en vigueur, viendra moderniser la Convention STE 108, d'où la référence commune à « 108+ »).

⁵⁴³ Article 14 de la Convention 108+.



textes et de données (*text and data mining* – TDM) sont largement utilisées pour le moissonnage et l'analyse de données. L'utilisation de contenus protégés par des droits d'auteur nécessite généralement l'autorisation des détenteurs de ces droits. Pour empêcher toute utilisation non autorisée, les organes de presse et les médias mettent en place des outils de protection contre le moissonnage de données, et les titulaires de droits d'auteur intentent des actions en justice contre les entreprises technologiques pour infraction, en faisant valoir que les résultats obtenus par l'IA s'appuient fortement sur des ensembles de données d'apprentissage susceptibles d'inclure leur matériel protégé par des droits d'auteur. Il apparaît ainsi que la réglementation ne s'étend pas encore spécifiquement aux questions liées à la propriété intellectuelle et que les détenteurs de droits ont l'impression de ne pas être protégés⁵⁴⁴.

La question de la légalité d'un recours non autorisé à du matériel protégé par le droit d'auteur pour entraîner des modèles d'IA relève également du Règlement sur l'IA. Selon le règlement, les fournisseurs de modèles d'IA générative sont tenus de mettre en place des mesures visant à respecter le droit de l'UE en matière de droit d'auteur et de droits voisins, ainsi qu'un résumé détaillé du contenu qui est utilisé pour entraîner le modèle d'IA générative⁵⁴⁵. Actuellement, le respect de la législation sur le droit d'auteur pose problème aux fournisseurs, à moins qu'ils n'aient obtenu une licence et, partant, une autorisation d'utilisation de la part du titulaire des droits. Il sera difficile de déterminer ce qui sera considéré comme un résumé suffisamment détaillé du contenu utilisé pour l'apprentissage du modèle d'IA générative jusqu'à ce que le Bureau de l'IA de l'UE⁵⁴⁶, récemment créé, fournisse un modèle pour ce type de résumé. Selon le considérant 107 du Règlement sur l'IA, ce résumé doit être généralement complet afin de permettre aux parties ayant des intérêts légitimes, y compris les titulaires de droits d'auteur, d'exercer et de faire respecter leurs droits en vertu de la législation de l'UE. Outre les aspects liés au droit d'auteur, l'obligation de divulgation est l'un des moyens de satisfaire au principe de transparence inscrit dans tous les instruments relatifs à l'IA mentionnés ci-dessus.

Le contenu généré par IA soulève également des questions sur la protection des droits d'auteur. Comme indiqué au chapitre 4.1.1, dans la plupart des juridictions, y compris aux États-Unis et dans l'UE, seule une création humaine peut bénéficier d'une protection par le droit d'auteur. Lorsque le contenu généré par ordinateur est protégé, la question se pose de savoir si la qualité d'auteur est simplement transférée à « l'humain derrière la machine » ou si les exigences jusqu'à présent élevées pour la protection en tant qu'« œuvre » sont réduites⁵⁴⁷. Une œuvre nécessite généralement un élément de créativité. On peut donc se demander si le produit d'une IA peut être considéré comme une nouvelle œuvre créative. Il faut tenir compte du fait que le résultat produit n'est pas

⁵⁴⁴ En ce qui concerne le manque de protection perçu, voir European Guild for Artificial Intelligence Regulation, *Our Manifesto for AI Companies Regulation in Europe* (4 novembre 2023).

⁵⁴⁵ Article 53(1)(c) et (d) du Règlement sur l'IA.

⁵⁴⁶ Le Bureau de l'IA de l'UE a été établi en mai 2024. Voir Commission européenne, Communiqué de presse, « La Commission crée un Bureau de l'IA afin de renforcer le rôle moteur de l'UE en matière d'intelligence artificielle sûre et digne de confiance » (29 mai 2024), https://ec.europa.eu/commission/presscorner/detail/fr/IP_24_2982. Voir également <https://digital-strategy.ec.europa.eu/fr/policies/ai-office-board>.

⁵⁴⁷ Voir chapitre 4.1.1.



simplement basé sur des modèles statistiques dans l'ensemble de données d'apprentissage et qu'il y a également une influence humaine significative dans le processus. Cette influence se manifeste lors de la création et de la configuration du système d'IA, ce qui inclut la sélection des données d'apprentissage, mais aussi dans la manière dont un système d'IA générative est guidé par les *prompts*, ainsi que par l'édition ultérieure des résultats générés par le système⁵⁴⁸. Comme le montre l'analyse de la jurisprudence au chapitre 4.1.2, le degré de contribution créative humaine qui doit également être reflété dans les contenus générés par l'IA varie d'un pays à l'autre et doit être évalué au cas par cas. Si le seuil est atteint, la qualité d'auteur est alors attribuée à l'utilisateur du système d'IA⁵⁴⁹.

La protection de la propriété intellectuelle ne se limite pas au droit d'auteur. Il existe des droits voisins qui ne requièrent pas de créativité mais protègent l'investissement ou les efforts économiques ou organisationnels consentis pour réaliser une œuvre. Dans le secteur audiovisuel, le droit voisin des producteurs de films en est un exemple. Comme ces droits ne dépendent pas d'un élément créatif, il est beaucoup plus facile de les faire valoir dans le cas d'un contenu généré par IA qui n'atteindrait pas le seuil requis pour bénéficier d'une protection au titre du droit d'auteur⁵⁵⁰.

Il est évident que les questions de droit d'auteur et de droits voisins posent de nouveaux défis dans le secteur audiovisuel, lorsque, par exemple, il n'est plus possible d'affirmer que le fournisseur d'effets spéciaux d'un film détient tous les droits sur le matériel fourni. Il est donc d'autant plus important d'assurer une transparence totale sur l'utilisation de l'IA. À cet égard, la convention-cadre et, dans certains cas, le Règlement sur l'IA exigent la transparence. Toutefois, il reste à voir si cela signifie également qu'un fournisseur de contenu doit divulguer l'utilisation de l'IA à d'autres parties de la chaîne d'approvisionnement dans le cadre de la production audiovisuelle. L'un des risques, notamment en ce qui concerne le contenu audiovisuel généré par l'IA, est que des éléments d'un travail antérieur contenus dans les données d'apprentissage puissent être reconnaissables dans le produit final. La détermination et les conséquences de la reconnaissabilité d'une œuvre antérieure est une question originale de droit d'auteur qui, encore une fois, n'est pas spécifique à l'IA. Des problèmes se posent lorsque l'utilisateur d'un contenu généré par l'IA ne sait pas qu'il utilise des œuvres de tiers et qu'il ne peut donc pas, du moins en vertu de la législation européenne, être tenu pour responsable du préjudice causé. En ce qui concerne le fournisseur d'IA, si l'on reprend l'exemple du droit de l'UE, ni la directive sur la responsabilité du fait des produits défectueux⁵⁵¹, ni la proposition de directive sur la responsabilité en matière d'IA⁵⁵², qui doit compléter le

⁵⁴⁸ Ces étapes ont été comparées par Baumann et Nordemann aux différentes étapes de la photographie de portrait, voir chapitre 4.1.2.

⁵⁴⁹ Voir chapitre 4.1.3.

⁵⁵⁰ Voir *ibid.*

⁵⁵¹ [Directive 85/374/CEE du Conseil du 25 juillet 1985 relative au rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de responsabilité du fait des produits défectueux](#), JO L 210/29, 7 août 1985.

⁵⁵² Commission européenne, [Proposition de directive du Parlement européen et du Conseil relative à l'adaptation des règles en matière de responsabilité civile extracontractuelle au domaine de l'intelligence artificielle \(Directive sur la responsabilité en matière d'IA\)](#), COM/2022/496 final.



Règlement sur l'IA, ne couvrent la question de la détermination de la responsabilité en cas de violation du droit d'auteur par des contenus générés par IA⁵⁵³. Cette incertitude juridique ne devrait pas manquer d'entraîner des litiges à l'avenir.

9.2.3. Aspects relatifs aux droits de la personnalité

Outre le risque de violation des droits d'auteur, les systèmes d'IA et, une fois de plus, les modèles d'IA générative en particulier, facilitent les atteintes aux droits de la personnalité lorsque le contenu usurpe l'identité d'une personne ou ses attributs, par exemple en générant des *deepfakes*, y compris des *deepfakes* basés uniquement sur la voix. Les droits de la personnalité permettent en effet à un individu de protéger et de contrôler l'utilisation de son image ou d'autres attributs personnels tels que sa voix⁵⁵⁴. Cette protection permet à l'individu d'exploiter économiquement l'utilisation de son identité. Le chapitre 5 énumère une série d'exemples d'utilisations non autorisées de l'image et de la voix, allant des images intimes non consenties générées par l'IA à l'amélioration de la voix d'un chanteur assistée par l'IA, en passant par le clonage de la voix. Il découle des aspects examinés concernant le droit d'auteur que des informations transparentes sur les données d'apprentissage, les fonctionnalités et les produits d'un système sont également cruciales pour le respect des droits de la personnalité, afin d'identifier les infractions et d'engager la responsabilité des contrevenants. La transparence est l'un des premiers principes fondamentaux énoncés initialement dans les documents d'orientation concernant l'IA, et figure également dans le Règlement sur l'IA et la convention-cadre⁵⁵⁵.

L'obligation de transparence inscrite dans le Règlement sur l'IA est toutefois limitée aux interactions directes avec un système d'IA, à moins que l'interaction avec celui-ci ressorte « clairement du point de vue d'une personne physique⁵⁵⁶ ». Lorsqu'il peut être moins évident qu'une personne est exposée à un contenu généré par l'IA, par exemple dans le cas de l'IA générative, les fournisseurs sont soumis à une obligation plus complète et doivent veiller à ce que les résultats générés par une IA générative soient marqués dans un format lisible par une machine et identifiables comme ayant été générés ou manipulés artificiellement⁵⁵⁷. Comme mentionné précédemment, cela soulève la question de savoir si l'utilisation de l'IA doit être divulguée tout au long de la chaîne d'approvisionnement d'une production audiovisuelle afin de permettre au déployeur de se conformer à son obligation de transparence énoncée à l'article 50 du Règlement sur l'IA. Lorsqu'un outil d'IA est utilisé pour créer des *deepfakes*, la divulgation est toujours

⁵⁵³ Voir chapitre 4.2.3.

⁵⁵⁴ Voir chapitre 5.1.

⁵⁵⁵ Voir sur la question de la transparence Cappello M. (éd.), *La transparence en matière de propriété des médias*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2021 et Cappello M. (éd.), *Transparence et responsabilité en matière d'algorithmes des services numériques*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2023.

⁵⁵⁶ Article 50(1) du Règlement sur l'IA.

⁵⁵⁷ Article 50(2) du Règlement sur l'IA.



obligatoire⁵⁵⁸. Les contrefaçons profondes faisant partie d'une « œuvre ou d'un programme manifestement artistique, créatif, satirique, fictif ou analogue » peuvent se limiter à divulguer le recours à l'IA « d'une manière appropriée qui n'entrave pas l'affichage ou la jouissance de l'œuvre ». Dans l'ensemble, les exigences de transparence et les questions relatives à l'utilisation de l'IA dans la phase de production constituent l'un des principaux défis pour le secteur de l'audiovisuel.

Il n'y a pas d'indications sur la manière dont un déployeur peut satisfaire à ces obligations de transparence en ce qui concerne les œuvres audiovisuelles, c'est-à-dire où, quand et de quelle manière les informations doivent être placées ou précisément formulées. À ce jour, on ne sait pas s'il suffira, par exemple, d'insérer une clause à cet effet dans le générique d'un film. Le rapport explicatif de la convention-cadre suggère, comme moyen de divulgation, l'étiquetage ou le filigrane pour permettre l'identification du contenu généré par l'IA. Toutefois, la transparence dans la convention-cadre vise davantage la préservation de la confiance du public, la protection des consommateurs et la prévention des ingérences électorales, tandis que la protection contre l'utilisation non autorisée des droits de la personnalité relève de l'interaction des principes de la dignité humaine, de l'autonomie individuelle et de la protection de la vie privée⁵⁵⁹. Contrairement à la législation émergente aux États-Unis qui cherche à protéger les individus contre le clonage et les *deepfakes* par IA non autorisés⁵⁶⁰, ni le Règlement sur l'IA ni la convention-cadre n'abordent précisément cette question en raison de leur complémentarité avec les cadres existants en matière de droits de l'homme. L'accord SAG-AFTRA décrit au chapitre 6 est le premier exemple d'accord d'un syndicat d'acteurs de cinéma et de télévision qui aborde également la dimension des droits de la personnalité.

9.2.4. La désinformation : un défi de taille

Il convient également de souligner les effets que peuvent avoir les *deepfakes* sur les destinataires. Tout d'abord, lorsque le public ne peut pas faire la distinction entre un vrai et un faux contenu, la crédibilité du fournisseur audiovisuel peut être ébranlée. Dans le même temps, des *deepfakes* de grande qualité peuvent avoir un impact significatif sur le discours public et influencer l'opinion publique⁵⁶¹. La valeur factuelle des images et du matériel vidéo est perdue lorsqu'il n'est plus possible de distinguer le contenu réel du contenu synthétique⁵⁶². En outre, les robots dotés d'IA sont capables de propager de fausses informations. Il faut aussi souligner que les systèmes d'IA ne produisent pas nécessairement des contenus reflétant la vérité, mais qu'ils peuvent aussi générer des informations erronées. Lorsque de telles informations sont intégrées à un support journalistique sans avoir été vérifiées, elles peuvent avoir les mêmes conséquences

⁵⁵⁸ Article 50(4) du Règlement sur l'IA.

⁵⁵⁹ Voir chapitre 5.3.2.

⁵⁶⁰ Pour une vue d'ensemble de la législation émergente et des projets de loi aux États-Unis, voir chapitre 5.4.

⁵⁶¹ Voir chapitre 7.2. Voir aussi Angelica Fernandez, « "Deep fakes" : disentangling terms in the proposed EU Artificial Intelligence Act », 85 (2) UFITA 2021, p. 392-433.

⁵⁶² Ruth Meyer et Rupprecht Podszun, *Künstliche Intelligenz und die Medienpolitik* (2024) ZRP 41.



négligentes que celles mentionnées plus haut : miner la confiance et la crédibilité du support et influencer le discours et l'opinion publics. Comme indiqué ci-dessus, en vertu du Règlement sur l'IA, les déployeurs d'un système d'IA qui produit des *deepfakes* doivent toujours faire preuve de transparence, mais il n'existe pas d'indications sur la manière dont cette transparence peut être assurée dans le secteur audiovisuel. En revanche, les fournisseurs d'un système d'IA destiné à interagir directement avec des personnes physiques ne doivent divulguer le recours à un système d'IA que lorsque celui-ci n'est pas évident. De même, les déployeurs de systèmes d'IA qui génèrent ou manipulent des textes qui sont publiés dans le but d'informer sur des questions d'intérêt public doivent indiquer que le texte a été généré ou manipulé artificiellement, sauf si le contenu a fait l'objet d'un processus d'examen humain ou d'un contrôle éditorial et si une personne physique détient la responsabilité éditoriale pour la publication du contenu⁵⁶³. En conséquence, des règles différentes s'appliquent au texte par rapport aux médias audiovisuels, et les fournisseurs de médias audiovisuels sont toujours tenus de divulguer le recours à un système d'IA.

En ce qui concerne la désinformation, il faut tenir compte du fait que des utilisateurs malveillants peuvent intentionnellement propager des documents manipulés. Alors que les services d'hébergement sont tenus de retirer les contenus illégaux en vertu du DSA⁵⁶⁴ et de la Directive sur le commerce électronique⁵⁶⁵, le Code de bonne pratique volontaire contre la désinformation⁵⁶⁶ (2022) considère que les comportements manipulateurs assistés par l'IA (par exemple, l'usurpation d'identité et les *deepfakes* malveillants) sont inadmissibles. Compte tenu des risques associés aux campagnes de désinformation générées par l'IA et à leur amplification algorithmique, la vérification des faits devient essentielle, de même que le renforcement de la confiance dans les médias en tant que gardiens de la vie publique. Cette tâche sera difficile si l'on considère que les systèmes de curation et de recommandation pilotés par l'IA parviennent à personnaliser le contenu, ce qui a pour effet d'enfermer plus ou moins les utilisateurs individuels dans une bulle où les opinions existantes sont renforcées et les opinions contradictoires, écartées⁵⁶⁷.

Le déploiement croissant des systèmes d'IA entraîne une transformation du travail qui n'est pas l'apanage du secteur des médias. Par conséquent, les initiatives portant sur les questions liées à l'emploi sont souvent de nature assez générale. Dans le cas du secteur audiovisuel, les travailleurs sont particulièrement susceptibles de perdre le contrôle sur leur travail et des problèmes se posent en ce qui concerne la paternité et la propriété des créations. Comme nous l'avons souligné dans les chapitres précédents, les

⁵⁶³ Article 50(4) du Règlement sur l'IA, considérant 134 du Règlement sur l'IA.

⁵⁶⁴ [Règlement \(UE\) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la Directive 2000/31/CE \(2022\) JO L 277/1.](#)

⁵⁶⁵ [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 \(2000\) JO L 178/1.](#)

⁵⁶⁶ [Code de bonnes pratiques renforcé contre la désinformation](#) (16 juin 2022).

⁵⁶⁷ Voir chapitre 8.2. Voir également Christina Etteldorf, « Les travaux normatifs du Conseil de l'Europe », in : Cappello M. (éd.), *Transparence et responsabilité en matière d'algorithmes des services numériques*, IRIS Spécial, Observatoire européen de l'audiovisuel, Strasbourg, 2023, p. 4, 5 et s.



technologies de l'IA ont par exemple été utilisées pour créer des « versions plus jeunes » d'acteurs ou pour faire revivre les voix de chanteurs décédés⁵⁶⁸. En outre, les modèles d'IA générative sont capables d'écrire des scénarios et de générer des clips vidéo. Alors que les préoccupations relatives à ces questions ont abouti à des mouvements de grève aux États-Unis et à la conclusion d'accords sectoriels avec les syndicats, le Parlement européen cherche à améliorer les conditions de travail des travailleurs du secteur audiovisuel en général et a adopté une résolution relative à un cadre de l'Union pour la situation sociale et professionnelle des artistes et des travailleurs des secteurs de la culture et de la création⁵⁶⁹. Dans le même temps, il est possible que des organisations de gestion collective, des associations et des fédérations européennes ou internationales, ainsi que des syndicats, deviennent plus actifs⁵⁷⁰.

9.2.5. Portée du Règlement sur l'IA : champ d'application géographique

Comme on peut le constater, le Règlement sur l'IA est le principal point de référence pour des règles détaillées relatives aux systèmes d'IA, en particulier au sein de l'UE. C'est pourquoi il a été largement mis en avant dans les chapitres précédents de cette publication et dans la présente synthèse. Néanmoins, au-delà des nouveautés apportées par ces règles, il est également nécessaire de mentionner brièvement le large champ d'application et, par conséquent, la portée territoriale du texte. Comme d'autres règlements relatifs à la sécurité des produits, le Règlement sur l'IA a des effets qui dépassent les frontières géographiques de l'UE. Il s'applique aux « fournisseurs établis ou situés dans l'Union ou dans un pays tiers qui mettent sur le marché ou mettent en service des systèmes d'IA ou qui mettent sur le marché des modèles d'IA à usage général dans l'Union⁵⁷¹ ». Il s'applique aux déployeurs qui ont leur lieu d'établissement ou qui sont situés dans l'UE⁵⁷² ; lorsque le contenu produit par un système d'IA est utilisé dans l'Union, mais que le fournisseur et le déployeur sont situés dans un pays tiers⁵⁷³ ; et aux importateurs et distributeurs de systèmes d'IA⁵⁷⁴.

Cela témoigne de l'ambition de permettre aux citoyens de l'Union – et à ses entreprises – d'utiliser les systèmes d'IA en toute sécurité et en toute confiance, quel que soit l'endroit où se trouvent les systèmes et leurs fournisseurs. C'est l'UE en tant que marché qui constitue le facteur de rattachement, de la même manière que pour le RGPD, pour ce qui est de l'applicabilité des règles de l'UE.

⁵⁶⁸ Voir chapitres 5.1., 6.2.3.3. et 10.4.

⁵⁶⁹ Parlement européen, [Cadre de l'Union pour la situation sociale et professionnelle des artistes et des travailleurs des secteurs de la culture et de la création](#), (2023/2051(INL)).

⁵⁷⁰ Voir chapitre 6.4.

⁵⁷¹ Article 2(1)(a) du Règlement sur l'IA.

⁵⁷² Article 2(1)(b) du Règlement sur l'IA.

⁵⁷³ Article 2(1)(c) du Règlement sur l'IA.

⁵⁷⁴ Article 2(1)(d) du Règlement sur l'IA.



9.3. Les perspectives d'avenir : « l'épreuve du temps » et les normes mondiales

9.3.1. Vers des normes globales et flexibles basées sur le risque dans une législation spécifique

Au niveau de l'UE, le Règlement sur l'IA, en tant que règlement portant sur la sécurité des produits, vise à réduire les risques pour l'homme lors de l'utilisation du produit « systèmes d'IA », mais il n'est pas certain qu'il apporte une véritable sécurité juridique sur ce nouveau marché. Certaines dispositions sont rédigées de manière très générale, ce qui veut dire qu'elles doivent être interprétées, voire précisées. En conséquence, plusieurs dispositions plutôt génériques doivent être clarifiées par des actes d'exécution de la Commission européenne. Par exemple, la Commission peut adopter des actes d'exécution établissant des spécifications communes pour les exigences relatives aux systèmes d'IA à haut risque⁵⁷⁵. Cette méthode de réglementation a été choisie, car un Règlement sur l'IA moins abstrait et plus détaillé aurait pu entraîner un risque d'obsolescence rapide des règles. Ainsi, le fait de « se contenter » d'énoncer des exigences normatives générales au niveau du droit dérivé de l'UE est caractéristique d'une réglementation technologique moderne, du moins lorsque la technologie évolue rapidement. Des orientations sur l'interprétation de certaines obligations peuvent être fournies par des normes techniques qui précisent les attentes et, par exemple dans le cas du Règlement sur l'IA, qui guident les fournisseurs tout au long du processus d'évaluation des risques. Si les fournisseurs bénéficient d'une présomption de conformité lorsqu'ils adhèrent à ces normes, leur inconvénient tient au fait qu'elles sont *de facto* établies par des acteurs privés et qu'il existe un risque que leur élaboration ne s'accompagne pas d'une plus grande transparence⁵⁷⁶.

Bien que beaucoup reste à faire en termes de mise en œuvre concrète de cette législation souple mais contraignante, il convient de prendre du recul et de saluer ce qui a déjà été réalisé, étant donné que la réglementation de l'IA n'est devenue pressante que ces dernières années, en raison du déploiement à grande échelle de produits et de services destinés au grand public. Au niveau général, on peut observer que les principes et les valeurs ont évolué jusqu'à faire l'objet d'un consensus et qu'ils sont souvent considérés comme la base principale (et minimale) attendue pour l'encadrement des systèmes d'IA. Ils sont désormais également inscrits dans le premier traité international

⁵⁷⁵ Article 41(1) du Règlement sur l'IA.

⁵⁷⁶ Voir Michael Veale et Frederik Zuiderveen Borgesius, « [Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach](#) » (2021) 22 *Computer Law Review International* 97. Au niveau de l'UE, le CEN et le CENELEC travaillent sur les normes.



juridiquement contraignant sur l'IA ouvert à la ratification. En outre, les définitions juridiques des systèmes d'IA se rapprochent de plus en plus⁵⁷⁷.

L'OCDE s'est engagée dans des activités empiriques et politiques sur l'IA afin de soutenir les débats politiques qui émergent depuis près d'une décennie, ce qui a abouti à la première norme intergouvernementale sur l'IA. Les instruments juridiques non contraignants tels que la recommandation de l'OCDE et la recommandation de l'UNESCO ont reflété un consensus minimal sous la forme de valeurs et de principes communs qui peuvent également être salués en tant que terrain d'entente ambitieux. Ce consensus est désormais également inscrit dans des instruments juridiquement contraignants, et la possibilité d'une approche plus globale a été notamment établie par la convention-cadre du Conseil de l'Europe. Non seulement la convention-cadre est ouverte à la signature de pays tiers, elle est également le produit d'une réflexion commune menée par un groupe de parties prenantes et d'experts internationaux et témoigne d'un travail collaboratif. En outre, la convention introduit une terminologie commune qui peut servir de base à une réglementation ultérieure, y compris, le cas échéant, à une réglementation sectorielle. Les limites, par exemple en ce qui concerne son applicabilité au secteur privé, tiennent au fait que l'instrument est un texte de compromis entre des États qui n'ont pas tous la même vision du niveau d'obligations qui devrait être imposé au secteur privé. Compte tenu de ces divergences peu surprenantes entre les approches des différents États, le fait qu'un tel accord ait été atteint constitue une avancée significative qui ouvre la voie à un niveau minimum plus global de règles contraignantes.

Outre le Règlement sur l'IA de l'UE qui a déjà été adopté, plusieurs pays du monde travaillent à l'élaboration d'une réglementation nationale en matière d'IA, en partie en réponse aux engagements pris à la suite des recommandations mentionnées plus haut. Les exemples présentés dans les chapitres précédents indiquent toutefois que la plupart des règlements sont limités à une certaine application (du système d'IA), par exemple le règlement sur l'IA du Tennessee (*Ensuring Likeness Voice and Image Security - ELVIS*) concernant le clonage indésirable de la voix, ou les projets de loi actuellement débattus au niveau fédéral aux États-Unis, tels que *Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES)*, qui vise à donner aux individus un droit de propriété sur leurs « répliques numériques », ou *No Artificial Intelligence Fake Replicas and Unauthorized Duplications (No AI FRAUD)*, qui va encore plus loin dans la protection de « l'image » d'une personne⁵⁷⁸.

⁵⁷⁷ Pour des définitions des systèmes d'IA, voir chapitre 1.1. Voir également Sandra Schmitz-Berndt, « AI Regulation and Governance on a Global Scale: Overview of Scope, Definitions and Key Elements », Annexe à Mark D. Cole, « AI Regulation and Governance on a Global Scale: An Overview of International, Regional and National Instruments » (2024) 1(1) *AIRe* 126, p. 141-142.

⁵⁷⁸ Voir à ce propos le chapitre 5.4.



9.3.2. Examen des aspects spécifiques au secteur des médias (audiovisuels)

Le Règlement sur l'IA se distingue par le fait qu'il s'agit d'un cadre complet, axé sur les produits. Avec son approche horizontale générale, il n'aborde pas les particularités de certains secteurs, mais tente de trouver des règles de base globales. La nécessité d'établir des règles différentes pour certains risques et de définir des orientations plus techniques a conduit à l'ajout de treize annexes au Règlement sur l'IA.

En général, les initiatives mondiales et régionales évitent d'aborder les caractéristiques propres à certains secteurs⁵⁷⁹. Là encore, il s'agit d'énoncer des principes généraux qui peuvent et doivent être appliqués universellement. Par conséquent, les questions relatives au secteur audiovisuel, en particulier les risques liés au pluralisme des médias et aux nouvelles formes de concentration du marché dues aux effets d'échelle, doivent être abordées dans une législation plus spécifique. Si le pluralisme des médias figure dans le Règlement sur la liberté des médias (EMFA) de l'UE⁵⁸⁰ récemment adopté, la question de la concentration du marché, qui est également visée par l'EMFA, est désormais à l'ordre du jour de la Commission européenne dans le contexte des systèmes d'IA. La Commission examine actuellement comment les géants de la technologie pourraient limiter le développement de modèles concurrents d'IA générative ou favoriser l'intégration de leurs applications d'IA dans d'autres produits et écosystèmes⁵⁸¹.

Même si les approches réglementaires adoptées jusqu'à présent ne concernent pas spécifiquement le secteur des médias ou de l'audiovisuel, l'importance des systèmes d'IA pour ce secteur est évidente. C'est pourquoi les fournisseurs de médias ont commencé à se positionner publiquement sur l'utilisation de l'IA dans la production et la diffusion de leurs contenus. Ces engagements répondent en partie aux obligations légales émergentes, mais déjà différents types de fournisseurs de médias, soit individuellement, soit par l'intermédiaire d'associations de journalistes ou d'éditeurs, se sont engagés dans une démarche d'autorégulation quant à la manière dont ils utilisent ou non les systèmes d'IA dans leurs processus. Parmi les exemples récents, citons le Conseil allemand de la presse qui a déjà modifié le « Pressekodex » (le code déontologique d'autorégulation⁵⁸²), le radiodiffuseur public allemand Bayerischer Rundfunk (BR) qui a mis à jour ses lignes directrices de 2020 en matière d'éthique vis-à-vis de l'IA à la lumière de l'émergence des

⁵⁷⁹ Voir Jo Pierson, Aphra Kerr *et al.*, « [Governing artificial intelligence in the media and communications sector](#) », (2023) 12(1) *Internet Policy Review*.

⁵⁸⁰ [Règlement \(UE\) 2024/1083 du Parlement européen et du Conseil du 11 avril 2024 établissant un cadre commun pour les services de médias dans le marché intérieur et modifiant la Directive 2010/13/UE \(Règlement européen sur la liberté des médias\)](#) (2024) JO L 1083/1.

⁵⁸¹ Commission européenne, [Concurrence dans l'IA générative et les mondes virtuels – appel à contributions](#) (9 janvier 2024).

⁵⁸² Le préambule précise désormais que toute personne qui s'engage à adhérer au Pressekodex porte la responsabilité éthique pour toutes ses contributions éditoriales, quelle que soit la manière dont elles sont créées, et que cette responsabilité s'étend aux contenus générés artificiellement. Voir aussi Presserat, Communiqué de presse, [Redaktionen auch für KI-generierte Inhalte ethisch verantwortlich](#), 18 septembre 2024.



GPAI⁵⁸³, la Sveriges Radio en Suède qui a introduit des principes relatifs aux GPAI⁵⁸⁴, le radiodiffuseur suisse SRG SSR qui a mis en œuvre des « principes nationaux en matière d'IA⁵⁸⁵ » ou le radiodiffuseur public britannique BBC qui a défini les principes qui déterminent l'approche de la BBC en matière de recours aux GPAI⁵⁸⁶.

Afin de mieux faire face aux défis que posent les systèmes d'IA aux médias – et plus concrètement au secteur audiovisuel – une première étape majeure consiste à s'engager sur la voie d'une plus grande transparence, comme l'exigent les règles mises en place jusqu'à présent. Par exemple, les obligations d'étiquetage permettent aux lecteurs, aux auditeurs et aux téléspectateurs d'être au moins conscients de cette nouvelle dimension de la production de contenu et/ou de l'exposition au contenu. Compte tenu de l'impact potentiel (négatif) sur les mécanismes de formation de l'opinion publique sur lesquels reposent les démocraties, la prochaine étape devra consister à évaluer plus avant si « l'intégrité des processus démocratiques et le respect de l'État de droit » devraient faire l'objet d'une plus grande attention. L'article 5 de la convention-cadre souligne la nécessité « d'adopter ou de maintenir des mesures visant à garantir que les systèmes d'intelligence artificielle ne sont pas utilisés pour porter atteinte à l'intégrité, à l'indépendance et à l'efficacité des institutions et processus démocratiques ». Cet article souligne en outre l'importance de protéger les processus démocratiques tout au long du cycle de vie de l'IA, en garantissant notamment un accès équitable au débat public et en permettant aux individus de se forger librement une opinion. Cela souligne l'importance d'examiner « les effets des systèmes d'IA et de leur utilisation sur les médias et leur rôle dans un débat pluraliste⁵⁸⁷ ».

⁵⁸³ Katharina Brunner, Rebecca Ciesielski, Philipp Gawlik *et al*, [Unsere KI-Richtlinien im Bayerischen Rundfunk](#) (BR, 12 juillet 2024).

⁵⁸⁴ Olle Zachrisson, Communiqué de presse, [Därför publicerar Sveriges Radio en policy för generativ AI](#), 7 juillet 2023.

⁵⁸⁵ SRF SSR, [Principes IA nationaux de SSR](#) (2023).

⁵⁸⁶ BBC, Communiqué de presse, [Generative AI at the BBC](#), 5 octobre 2023.

⁵⁸⁷ Voir également la convention-cadre, [Rapport explicatif](#), paragraphe 46.



10. Dilemmes éthiques et défis sociétaux posés par l'IA générative

*Bart van der Sloot, Tilburg Institute for Law, Technology and Society de la Tilburg University*⁵⁸⁸

10.1. Introduction

Les précédents chapitres de cette publication ont présenté au lecteur un aperçu de l'IA, de ses rouages internes et des technologies de données qui l'alimentent. Ils se penchent en particulier sur l'utilisation de l'IA dans le secteur audiovisuel et sur les problématiques qui en résultent en termes de protection de la vie privée et des données, de droit d'auteur et de propriété intellectuelle, de responsabilité et d'obligation de rendre compte, de droits de la personnalité, de droit du travail, de désinformation et de pluralisme. De plus, le chapitre 9 nous offre une incursion dans le monde de demain avec des exemples des nombreuses répercussions de l'IA dans le secteur audiovisuel. Le présent chapitre conclut par une analyse approfondie des fondements éthiques dans lesquels sont ancrés non seulement nos concepts juridiques, mais aussi la société dans son ensemble. Nous examinerons en particulier dans quelle mesure ces fondements éthiques peuvent être ébranlés par les développements technologiques à l'œuvre, notamment par les avancées de l'IA générative (GenAI).

La deuxième partie du présent chapitre aborde brièvement les trois piliers essentiels de toute société ouverte : la confiance, la vérité et la confrontation. La troisième partie analyse comment l'IA affecte ces valeurs, en soulignant en particulier la perte d'authenticité, l'impact de la propagation de contenus manipulés et générés par l'IA sur la confiance, et la capacité de l'IA à décharger les gens de nombreuses tâches difficiles de la vie courante. La quatrième partie aborde la question éthique au regard du secteur audiovisuel et analyse les défis posés par la GenAI pour le journalisme et les médias d'information, s'interroge sur l'utilisation de l'IA dans l'industrie cinématographique et, enfin, se penche sur la différence entre la créativité de l'IA et la créativité humaine. La cinquième partie traite plusieurs défis sociétaux, en particulier en ce qui concerne la légitimité du processus démocratique, les effets de la GenAI sur les

⁵⁸⁸ Bart van der Sloot est maître de conférences au Tilburg Institute for Law, Technology and Society de la Tilburg University. Il a traité la question de l'IA générative (GenAI) de façon approfondie dans un ouvrage en libre accès : Van der Sloot, B., *Regulating the Synthetic Society: Generative AI, Legal Questions, and Societal Challenges*, Bloomsbury, 2024.



procédures juridiques, et les questions relatives à l'octroi d'une personnalité juridique aux entités pilotées par l'IA. En conclusion, la sixième partie présente une synthèse de ce qui précède.

10.2. Nos fondements éthiques

Ce paragraphe se penche sur le rôle majeur de trois fondements éthiques dans notre société contemporaine : la vérité, la confiance et la confrontation.

En premier lieu, il est difficile d'évaluer à quel point la vérité est essentielle pour les intérêts personnels, sociaux et sociétaux. « Connais-toi-même » est sans doute l'épigraphe le plus emblématique de la civilisation occidentale, qui conçoit, tout au moins en partie, la vie comme une quête permanente visant à se comprendre soi-même. Se tromper sur soi-même ou agir d'une manière qui semble déloyale envers soi-même peut être l'une des expériences personnelles les plus angoissantes. Dans les relations sociales, le fait d'avoir une conception commune de la vérité est nécessaire pour créer un lien, et bon nombre des institutions de la société reposent sur un contrat de base consacrant un certain nombre de principes fondamentaux tels que, notamment, la démocratie. Il n'est donc pas étonnant que dans les sociétés où il existe un profond désaccord sur les principes fondamentaux, comme aux États-Unis, le processus démocratique se fissure et que le lien social intercommunautaire se délite⁵⁸⁹.

Par ailleurs, la vérité est intrinsèquement liée à la notion de confiance. La confiance est fondamentale dans la perception qu'ont les gens de la réalité. Les gens sont confiants dans le fait que demain, le soleil se lèvera, leur ami sera toujours leur ami, et qu'ils seront toujours en mesure de marcher. Même si des changements se produisent, ils surviennent en général progressivement. Cette stabilité épistémologique sert de base à la perception qu'ont les gens du monde, des autres et d'eux-mêmes. À titre d'exemple, si une personne n'est pas sûre d'être vivante ou morte le lendemain, cela aura d'énormes répercussions sur sa capacité à interagir avec autrui et à trouver un sens à la vie⁵⁹⁰.

Enfin, en dernier lieu, la confiance et la vérité sont donc fondamentales d'une part, pour nouer des relations sociales, mener une réflexion sur soi-même et pouvoir évoluer, et d'autre part, pour les institutions de la société⁵⁹¹. Mais en dépit de cela, nous vivons de nombreuses situations où nous constatons que nous avons tort sur nous-mêmes, sur les autres ou sur le monde qui nous entoure, où il est impossible de se mettre d'accord sur une vision commune de la vérité, où notre confiance a été trahie, où nos prévisions

⁵⁸⁹ Baumer D. C., Gold H. J., *Parties, polarization and Democracy in the United States*, Routledge, 2015. Pausch M., « The future of polarisation in Europe: relative cosmopolitanism and democracy », *European Journal of futures Research* (2021), 9(1), p. 12. Horonziak S., « Dysfunctional democracy and political polarisation: the case of Poland », *Zeitschrift für Vergleichende Politikwissenschaft* (2022), 16(2), 265-289.

⁵⁹⁰ Keymolen E., *Trust on the line: a philosophical exploration of trust in the networked era*, Wolf Legal Publishers, 2016.

⁵⁹¹ Hegel G. W. F., *Phänomenologie des Geistes*, Suhrkamp, Francfort 1969. En particulier dans l'interprétation de Hegel par Kojève : Kojève A., *Introduction to the Reading of Hegel*, Cornell University Press, 1980.



probabilistes de l'avenir (proche) s'avèrent erronées. Alors que ces situations sont stressantes, voire même traumatisantes, et que notre seule envie est de les éviter, de les prévenir et d'y échapper, elles sont néanmoins indispensables à notre développement personnel. Même si nous préférions ne pas être confrontés à des conceptions du monde qui sont radicalement opposées aux nôtres, ne pas être confrontés aux derniers moments d'un proche atteint d'une maladie incurable, ou éviter un ami qui nous critique sans concession, c'est justement par la confrontation, par ces expériences pénibles et désagréables, que nous approfondissons notre connaissance de nous-mêmes, de nos proches et du monde qui nous entoure. Sans ces moments-là, nous serions des êtres statiques et unidimensionnels.

10.3. Impact de l'IA sur ces fondements

Ce paragraphe étudie certains effets potentiels de l'IA générative sur la vérité, la confiance et la confrontation.

Le premier défi est lié au fait que la généralisation des *chatbots*, des robots humanoïdes, de la réalité augmentée, de la réalité virtuelle, des hypertrucages et d'autres applications alimentées par la GenAI entraîne une augmentation considérable des contenus en ligne produits ou manipulés par l'IA. La GenAI n'est pas seulement utilisée pour induire délibérément les gens en erreur, mais aussi pour créer, entre autres, de belles images de cabanes en forêt artificielles sur Facebook ou des photos d'enfants fictifs dans des postures spectaculaires, qu'il est impossible de différencier d'un contenu authentique. Les appareils photo fonctionnent de plus en plus sur la base de concepts préprogrammés : une forêt en feu peut donc être encore verte sur la photo parce que l'IA « sait » qu'une forêt est verte et la lune peut briller d'un bel éclat même par une soirée brumeuse⁵⁹². Lors des visio-conférences, la configuration par défaut filtre les fréquences sonores les plus aigües et adoucit le teint des gens. La technologie de l'hypertrucage étant librement accessible à tous, les gens produisent de plus en plus d'hypertrucages à des fins humoristiques ou satiriques. Globalement, toutes ces tendances (et bien d'autres encore) peuvent signifier que dans quelques années, plus de 90 % de tous les contenus numériques seront générés ou manipulés par l'IA⁵⁹³, ce qui impliquerait l'avènement d'une ère post-vérité.

Le deuxième défi, lié à un impact plus structurel de la surproduction de contenus générés par l'IA, réside dans le fait que les gens ne sauront plus ce qu'ils doivent croire. On sait que les gens qui ont déjà cru, à tort, à un contenu mensonger sont, par la suite, plus circonspects lorsqu'ils voient des informations choquantes ou sensationnelles. L'incertitude quant à la véracité de la communication dans un monde presque entièrement numérisé et médiatisé peut avoir un impact déterminant, tant au niveau personnel que sociétal. La perte de confiance que connaissent déjà les « médias traditionnels » pourrait s'accélérer, et la tendance selon laquelle les gens choisissent les

⁵⁹² <https://www.theverge.com/2023/3/13/23637401/samsung-fake-moon-photos-ai-galaxy-s21-s23-ultra>.

⁵⁹³ Schick N., *Deepfakes and the infocalypse: what you urgently need to know*, Hachette, Royaume-Uni, 2020.



médias qui confirment leur vision préétablie du monde pourrait s'accroître. Cela peut impliquer que les différents groupes deviennent de plus en plus focalisés sur leur propre perception de la réalité, ce qui peut conduire à une polarisation et à un malaise au sein de la société. Une société synthétique peut donc provoquer une insécurité épistémologique : ce que je vois est-il réellement vrai ? Est-ce que je parle réellement à la personne à qui je crois parler ? Est-ce une personne humaine ou un robot humanoïde ? L'avatar qui ressemble à un ami est-il réellement contrôlé par cet ami ? L'avatar est-il une « version encore plus authentique » de la personne physique qui le contrôle ou une version trompeuse, ou bien l'avatar physique et l'avatar virtuel sont-ils tous deux des représentations tout aussi importantes d'une personne ? L'avatar est-il généré et/ou contrôlé par l'être humain ou par l'IA ? Sous l'effet conjugué d'une forme d'hyperpersonnalisation, où chacun vit dans sa propre réalité, et de la « séquestration de l'expérience »⁵⁹⁴, il peut être difficile de vérifier si ce que l'on a vécu est vrai : « L'hyperpersonnalisation peut donc entraîner une perte de confiance dans la perception (objective). Quelle valeur pouvons-nous accorder à une histoire ou à un témoignage si nous ne pouvons pas juger dans quelle mesure ce que la personne a vécu est "réel" ? »⁵⁹⁵.

Enfin, lorsque le travail domestique, en usine, voire même dans le domaine créatif est assuré par l'IA, les humains risquent de devenir de plus en plus dépendants d'entités pilotées par l'IA pour leurs activités physiques, intellectuelles et créatives. Bon nombre de gens, même si ce n'est pas tous, choisiront spontanément la voie de la facilité. Si un robot peut porter un sac à provisions, pourquoi le faire nous-mêmes ? Si un avatar virtuel peut emmener une personne visiter un pays étranger et que Google Lens assure la traduction des textes, pourquoi essayer d'apprendre les rudiments de cette langue en vacances ? Si un robot humanoïde peut faire office de partenaire amoureux, à quoi bon se rendre à un rendez-vous galant en prenant la peine de se mettre sur son trente-et-un ? Si un robot de soins à la personne peut aider un senior à aller aux toilettes, pourquoi ses enfants devraient-ils le faire ? Sachant que ChatGPT peut écrire une dissertation, pourquoi le faire soi-même ? De plus, grâce à la technologie de l'hypertrucage, il est possible de créer des bulletins d'information qui correspondent à la vision du monde de tout un chacun, il est également possible de ramener à la vie un partenaire décédé, de sorte que le survivant n'ait pas à être confronté au deuil et au chagrin ; les adolescents en surpoids victimes de harcèlement à l'école peuvent s'évader dans une réalité virtuelle avec un avatar qui correspond à leur idéal corporel, de sorte qu'ils n'ont pas à être confrontés au problème douloureux de leur aspect physique réel. Ainsi, à bien des égards, la GenAI a pour effet de faire disparaître toute confrontation de la vie des gens.

⁵⁹⁴ Giddens A., *Modernity and self-identity: self and society in the late modern age*, Stanford University Press, 1991.

⁵⁹⁵ Schermer Bart W. et Ham J. v., *Regulering van immersieve technologieën*, 2021.



10.4. Dilemmes éthiques dans le secteur audiovisuel

Ce paragraphe se penche sur trois dilemmes éthiques propres au secteur audiovisuel, à savoir l'impact de la GenAI sur le journalisme, son utilisation par l'industrie cinématographique, et la question plus générale de la créativité humaine par rapport à celle de l'IA.

Tout d'abord, il est clair que l'essor des médias synthétiques aura un impact sur le fonctionnement des organes d'information. Ils seront utilisés pour des applications positives. Par exemple, les technologies synthétiques peuvent être utilisées pour visualiser des situations où il n'est pas possible de filmer, comme dans les zones de guerre. Par ailleurs, l'IA peut générer des présentateurs ou présentatrices capables de parler n'importe quelle langue au monde. Ou encore, les médias peuvent améliorer l'expérience utilisateur en permettant le reportage participatif en 3D en direct et en réalité virtuelle. Mais cela peut aussi avoir des effets négatifs. Les médias ont déjà du mal à vérifier correctement l'exactitude et l'authenticité de l'ensemble des contenus générés par les utilisateurs et des publications en ligne. Prenons un petit exemple, qui n'en est pas moins édifiant : un footballeur a été sorti du terrain lors d'un match de championnat européen en raison d'un problème de santé, et peu de temps après une photo est apparue sur Twitter (rebaptisé X) pour prouver qu'il était encore en vie pendant son transport à l'hôpital. Cependant, il a fallu un certain temps à de nombreux médias traditionnels pour faire référence à cette photo, car elle aurait pu être fictive ou postdatée. Dans un monde où chaque individu a accès aux technologies synthétiques et peut créer et diffuser en quelques minutes des vidéos, photos ou séquences audio hypertruquées, la question est de savoir comment les médias peuvent garantir concrètement que leurs reportages restent authentiques. Les médias de qualité qui investissent dans ce type de procédures ne courent pas seulement le risque de réduire leurs revenus en raison du coût que cela implique, mais aussi de devenir « obsolètes », car d'autres médias, moins soucieux de diligence, sont plus rapides à couvrir un événement et publient sans hésiter des reportages à sensation, même si ceux-ci sont démentis ultérieurement. De plus, l'utilisation de la GenAI dans le journalisme soulève un certain nombre de problèmes : les outils de l'IA générative sont entraînés avec des données comportant des contenus des médias de service public (MSP). Si le MSP l'accepte, cela garantit aux citoyens l'accès à un contenu fiable⁵⁹⁶. Cependant, certains MSP peuvent ne pas vouloir traiter avec des entreprises d'IA. Cela doit-il être considéré comme faisant partie intégrante du mandat ou de la mission des MSP ? Cela pourrait changer la vision de ce qu'est le journalisme. Le débat pourrait prendre une dimension plus large et plus philosophique : comment préserver la touche d'humanité dans le journalisme ? Comment préserver le pluralisme des médias ? La réalité devrait rester soumise à une médiation humaine et non pas tant technologique. Que se passera-t-il si la culture dans son ensemble n'est plus créée et transmise par les êtres humains ?

⁵⁹⁶ Plus d'informations sur le site de l'UER : <https://www.ebu.ch/guides/loginonly/report/ai-regulation-and-its-importance-for-public-service-media-a-look-ahead>



Deuxième point, la GenAI est déjà utilisée dans l'industrie cinématographique, et son usage s'intensifiera à mesure que la technologie progresse. L'IA est utilisée, entre autres, pour rajeunir ou vieillir artificiellement les interprètes, pour ramener des interprètes décédés afin de terminer un film, ou pour réaliser des cascades dangereuses, tandis que dans le secteur des films réservés aux adultes, les actrices peuvent avoir une doublure créée par l'IA pour jouer les scènes les plus obscènes ou répondre aux demandes spéciales de leurs utilisateurs. Cela soulève la question suivante : si le cinéma devient généré par l'IA, le public continuera-t-il à croire aux histoires ? Les films ne seront-ils pas trop parfaits pour que le public puisse s'identifier aux personnages des contenus audiovisuels générés par l'IA ? À l'avenir, l'IA « risquant » de rendre les contenus parfaits, cela signifie-t-il que l'utilisation de la créativité humaine peut rendre les contenus imparfaits ? Mais la GenAI peut aussi avoir, entre autres, trois conséquences : tout d'abord, elle pourrait conduire au remplacement des êtres humains. Les scénaristes, par exemple, craignent que leur métier ne soit menacé par la capacité de la GenAI à produire non seulement de nouvelles idées, mais aussi des scénarios entiers, des descriptions de personnages et des épisodes de séries. Ensuite, les interprètes pourraient être remplacés par des versions d'eux-mêmes ou de personnes fictives produites par l'IA, ce qui, pour l'industrie, aurait l'avantage non seulement de réduire les coûts, mais d'offrir une plus grande flexibilité. Il sera rapide et peu coûteux, par exemple, de produire un film et de le montrer à un public test, puis de refaire quelques scènes ou reprendre le scénario. Enfin, ce qui est nouveau avec la GenAI, c'est qu'elle s'est démocratisée. Elle donne à chacun les outils pour faire un film à son goût. Non seulement cela alimente la crainte du remplacement des êtres humains dans le monde du travail, mais surtout, les particuliers ne sont pas soumis à des normes professionnelles strictes comme le sont les industries en place. Pour ne citer qu'un exemple parmi d'autres, selon certains rapports, plus de 95 % de l'ensemble des hypertrucages concernent des scènes pornographiques imposant des actes sexuels à des femmes sans leur consentement⁵⁹⁷. Les femmes célèbres et les femmes politiques sont régulièrement prises pour cible, ce qui mine leur crédibilité et nuit à leur réputation, et a conduit certaines d'entre elles à démissionner pour se protéger, elles-mêmes et leurs proches, contre des contenus préjudiciables. Autres conséquences, sans doute encore plus problématiques, de la démocratisation et de l'accessibilité de la technologie de l'hypertrucage : le corps féminin est hypersexualisé, les idéaux de beauté irréalistes sont renforcés, les femmes sont stigmatisées, et n'importe quel adolescent peut générer une vidéo porno-synthétique d'une fille de sa classe et la diffuser en privé, sur les réseaux sociaux ou sur des sites porno spécialisés. Cela peut avoir des effets dévastateurs sur le statut social des jeunes filles, leur image de soi et leur développement personnel.

De plus, la GenAI soulève des questions complexes en ce qui concerne la créativité et la propriété intellectuelle. Par exemple, la plupart des modèles d'IA sont entraînés en moissonnant tout simplement le plus de données possible sur Internet. Or, il s'agit de savoir si cette pratique relève de l'exception relative à l'utilisation équitable, si cette information doit être considérée comme libre d'utilisation, ou si cette pratique constitue une violation flagrante des droits de propriété intellectuelle et de la possibilité

⁵⁹⁷ https://regmedia.co.uk/2019/10/08/deepfake_report.pdf



pour les créatifs de percevoir une rémunération pour leurs productions. De plus, il n'existe pas de règles claires en matière de droits de propriété intellectuelle sur les contenus produits par l'IA. Supposons qu'une personne demande à un grand modèle de langage (LLM) de regarder tous les films des frères Coen et de produire pour elle un nouveau film dans le plus pur style Coen. Qui a les droits de propriété intellectuelle sur ce nouveau film : les frères Coen, la personne qui a émis la demande, la société d'IA, aucun d'entre eux, ou tous à la fois ? La créativité générée par l'IA pose également la question suivante : qu'est-ce que la créativité ? Qu'est-ce qu'une nouvelle œuvre, qu'est-ce qu'il faut définir comme étant le fruit d'un processus intellectuel, et l'IA pourra-t-elle jamais être créative à la façon des humains⁵⁹⁸ ? Aucune de ces questions ne peut recevoir de réponse appropriée, mais à terme, les paradigmes juridiques devront clarifier les choses et les choix qui seront faits seront déterminants pour l'avenir de la créativité.

10.5. Défis sociétaux posés par l'IA

Ce paragraphe se penche sur trois défis sociétaux soulevés par la GenAI : l'intégrité électorale, la validité des preuves présentées dans les procédures judiciaires, et la question de la sensibilité de l'IA ainsi que la nécessité ou non de protéger à terme les entités dirigées par l'IA.

Le premier défi concerne l'utilisation de contenus basés sur l'IA à des fins politiques. Ces technologies sont utilisées par les personnalités politiques, les partis politiques, les organisations qui les soutiennent et leurs partisans, ainsi que par leurs adversaires. À titre d'exemples, certains responsables politiques ont leur hologramme en tournée nationale⁵⁹⁹, créent des hypertrucages d'eux-mêmes pour s'adresser à leur électorat pendant leur incarcération⁶⁰⁰, font retoucher leur visage pour paraître plus sympathiques⁶⁰¹, ou utilisent la technologie de l'hypertrucage pour prononcer un discours dans une langue minoritaire qu'ils ne parlent pas⁶⁰². Alors que ces utilisations vont à l'encontre des conditions générales des divers produits et services, aucune disposition légale n'établit explicitement si elles constituent ou non une fraude électorale. La technologie de l'hypertrucage peut également être utilisée contre les personnalités politiques, par exemple en utilisant le clonage vocal pour faire tenir à un candidat des propos compromettants ou préjudiciables⁶⁰³ ou en créant des univers médiatiques entièrement factices : vidéo hypertruquée d'un candidat politique ayant un comportement scandaleux, sites de désinformation en ligne qui semblent rendre compte de l'incident, faux comptes X qui discutent de la vidéo hypertruquée, faux comptes Instagram qui

⁵⁹⁸ Voir chapitre 4 sur la protection des éléments produits.

⁵⁹⁹ <https://www.theverge.com/2014/5/7/5691714/indian-politician-uses-holograms-to-reach-voters>

⁶⁰⁰ <https://www.forbes.com/sites/siladityaray/2023/12/18/imran-khan-pakistans-jailed-ex-leader-uses-ai-deepfake-to-address-online-election-rally/>

⁶⁰¹ <https://www.cfr.org/blog/ai-context-indonesian-elections-challenge-genai-policies>

⁶⁰² <https://www.theverge.com/2020/2/18/21142782/india-politician-deepfakes-ai-elections>

⁶⁰³ <https://www.nbcnews.com/politics/2024-election/democratic-operative-admits-commissioning-fake-biden-robocall-used-ai-rcna140402>



gènèrent des mèmes en utilisant des captures de la vidéo, etc. L'utilisation préjudiciable de l'IA par les adversaires politiques est pratiquée à la fois par des adversaires politiques au niveau national et par des puissances étrangères. Certains pays comme la Russie, la Chine et l'Iran ciblent également les pays du Sud avec des informations fausses et manipulées pour diverses raisons, notamment pour peser sur une décision concrète (par ex. pour qu'une entreprise d'État russe obtienne un contrat), influencer les élections locales (par ex. pour amener au pouvoir un régime favorable à la Chine) ou influencer les décisions au niveau international (par ex. en incitant un pays à voter en faveur de la levée des sanctions contre l'Iran)⁶⁰⁴. Par conséquent, la technologie de l'hypertrucage comporte le risque d'être utilisée pour déstabiliser ou influencer le débat et la vie politiques.

Le deuxième défi est lié au fait que la généralisation des *chatbots*, des robots humanoïdes, de la réalité augmentée, de la réalité virtuelle, des hypertrucages et d'autres applications alimentées par la GenAI entraîne une augmentation considérable des contenus en ligne produits ou manipulés par l'IA. Actuellement, dans la plupart des juridictions, les moyens de preuve présentés dans une procédure judiciaire sont réputés authentiques sauf preuve du contraire. Or, cette présomption d'authenticité devra sans doute être remise en cause, puisque toute preuve fournie aura probablement été mise en contact avec l'IA et qu'il conviendra désormais d'établir qui a effectué quelles manipulations, dans quel but, et quelle est leur pertinence dans l'affaire en cours⁶⁰⁵. Les programmes de détection de l'IA ne peuvent filtrer que la moitié des contenus générés ou manipulés par l'IA et ne donnent souvent qu'une « probabilité d'authenticité » (par ex. « la probabilité que ce contenu soit authentique est de 67 % »). Il est donc clair que l'essor de la GenAI et, en particulier, la démocratisation de la technologie de l'hypertrucage, auront des répercussions importantes sur les procédures judiciaires. Il se peut que les procédures prennent plus de temps puisque les parties pourront contester l'authenticité de la preuve, il pourrait y avoir une plus grande dépendance à l'égard des experts, et il y aura une incertitude quant aux normes ou aux interdictions qui devraient être respectées par chacune des parties à telle ou telle étape de la procédure. Avec l'essor des contenus synthétiques, deux arguments ne manqueront pas d'être de plus en plus fréquemment invoqués en audience, à savoir : « À l'époque, je pensais que c'était vrai » (même si plus tard, il pourrait s'avérer que cette conviction était basée sur un hypertrucage) et « À l'époque, je pensais que ce n'était pas vrai » (*ibid.*). Aucun de ces arguments ne peut être exclu *a priori*. En outre, en l'absence de garanties adéquates, les juges seront de plus en plus amenés à commettre des erreurs en présumant à tort que des preuves sont authentiques ou non.

Le troisième point est lié à une question juridique qui se profile à l'horizon et qui ne sera pertinente, pour autant qu'elle le devienne un jour, que dans une décennie, voire

⁶⁰⁴ Dias J. A., Doca H. H. et da Silva F. F., « Bots, fake news, fake faces and deep fakes: automation, under the bias of dromology, as a sophisticated form of biopower to influence the democratic election process », *Revista de Ciências Jurídicas* (2021), 26(3), 1-14. Barari S., Lucas C. et Munger K., *Political deep fake videos misinform the public, but no more than other fake media*, OSF Preprints (2021), 13. Wilkerson L., « Still waters run deep (fakes): the rising concerns of "deepfake" technology and its influence on democracy and the first amendment Mo », *L. Rev.* (2021), 86, 407. Whyte C. (2020), « Deep fake news: AI-enabled disinformation as a multi-level public policy challenge », *Journal of cyber policy*, 5(2), 199-217.

⁶⁰⁵ <https://cyfor.co.uk/deepfake-audio-evidence-used-in-uk-court-to-discredit-father>.



un siècle : il s'agit de l'évolution de l'IA vers la sensibilité et l'intelligence générale artificielle (AGI). L'AGI désigne le stade où l'intelligence artificielle est équivalente à l'intelligence humaine, dont elle ne peut être différenciée. Il est clair que sur certains points, l'IA a déjà supplanté l'intelligence humaine, alors que dans d'autres domaines, elle est toujours à la traîne. Les capacités de l'IA ont progressé de façon exponentielle au cours des trois dernières années, et si elles continuent au même rythme pendant une autre décennie, l'AGI pourrait effectivement être atteinte⁶⁰⁶. En revanche, il est plus délicat d'établir l'existence d'une conscience de soi ou d'une sensibilité de l'IA. Cependant, ces deux aspects ne sont pas tout noirs ou tout blancs. Au fur et à mesure que la technologie progresse, l'IA sera capable d'effectuer des tâches toujours plus diversifiées tout en devenant de plus en plus intelligente. Même si l'IA ne remplace pas l'intelligence humaine dans tous les domaines de la vie, il est probable que dans de nombreux secteurs, l'IA deviendra plus performante que les humains. La question de la sensibilité dépend, en définitive, d'une expérience subjective, ce qui ne permet pas de vérifier objectivement ce qu'une entité ressent « réellement ». Indépendamment de ce qui se passe « réellement » à l'intérieur d'une machine, il est clair que les humains s'attachent facilement aux objets et leur attribuent volontiers des qualités anthropomorphiques. À titre d'exemple, le Tamagotchi, un ordinateur en forme d'œuf censé représenter un animal de compagnie virtuel, a montré à quelle vitesse les gens pouvaient développer un lien émotionnel. Ce constat ne peut que se renforcer avec les progrès actuels des robots mous, des LLM et de l'IA.

La question se pose donc de savoir quand et dans quelle mesure les entités dirigées par l'IA se verront octroyer à la fois une protection morale et juridique, voire même des droits « humains ». Par exemple, le fait d'avoir des machines intelligentes qui exécutent des tâches de production de longue haleine et dégradantes a été comparé à des formes modernes d'esclavage, tout comme le fait d'avoir des entités intelligentes faisant office de robots sexuels est parfois assimilé au proxénétisme. Un certain nombre de questions d'ordre moral et juridique se posent en ce qui concerne le statut des hypertrucages représentant des personnes fictives. Il est possible de produire des images et des vidéos très réalistes montrant des gens dont on jurerait qu'ils existent, alors qu'ils n'existent pas. Quelles barrières morales faudrait-il, le cas échéant, imposer à l'utilisation de l'image de personnes fictives ? Dans de nombreuses juridictions, mais certainement pas toutes, la production de pédopornographie montrant des personnes fictives est interdite, mais qu'en est-il de la sado-pornographie ultra-violente et avilissante représentant des personnes fictives ? Enfin, il y a la question des hypertrucages représentant des personnes décédées. Il n'existe actuellement aucune protection juridique explicite contre ce type d'utilisation, car les droits ne sont généralement attachés qu'aux personnes physiques vivantes⁶⁰⁷. À cet égard, il convient d'établir s'il appartient aux proches de la personne décédée de disposer de ses données et de son image, et s'il devrait y avoir une disposition légale par défaut qui interdise une telle option à moins que la personne concernée n'ait signifié dans son testament son accord

⁶⁰⁶ Voir toutefois Shumailov I., Shumaylov Z., Zhao Y., Gal Y., Papernot N. et Anderson R., *The curse of recursion: training on generated data makes models forget*, 2023.

⁶⁰⁷ Voir chapitre 5 sur les droits de la personnalité.



avec une telle utilisation ou, inversement, qui autorise cette utilisation à moins que les personnes concernées n'aient explicitement notifié leur refus d'une telle pratique.

10.6. Conclusion

Force est de constater que la GenAI soulève un grand nombre de dilemmes éthiques et de défis sociétaux. Le présent chapitre s'est penché sur les trois aspects les plus problématiques, à savoir l'évolution potentielle vers une société post-vérité, l'impact de la GenAI sur la confiance, et la disparition des situations de confrontation, c'est-à-dire les moments qui nous permettent, en dépit de leur caractère éprouvant, d'apprendre et d'évoluer dans la vie. Nous avons analysé en particulier les effets spécifiques de la GenAI sur le secteur audiovisuel et l'utilisation de la GenAI par les médias, mais aussi les défis que pose cette technologie en termes de vérification des contenus ; nous avons examiné l'utilisation de l'IA dans l'industrie cinématographique aux fins de décharger les interprètes et les scénaristes des tâches les plus fastidieuses, tout en considérant son impact potentiel sur la suppression de certains métiers humains et le fait que les utilisateurs ayant accès à cette technologie ne sont pas soumis aux mêmes normes éthiques que les professionnels ; nous avons également abordé la question de la propriété intellectuelle et de la propriété des contenus produits par l'IA, ainsi que la question de ce qu'est réellement la créativité. Enfin, nous avons passé en revue les défis de plus grande envergure pour la société en mettant l'accent sur trois aspects en particulier : la menace que représente la GenAI pour les élections démocratiques, ses effets sur les procédures judiciaires et l'état de droit, et la question de savoir si, tôt ou tard, les entités dirigées par l'IA pourront bénéficier d'une protection et s'il convient de fixer des barrières morales et juridiques à leur création et à leur exploitation.

Même s'il est clairement établi que des intérêts majeurs sont en jeu, tant au niveau individuel que sociétal, il n'y a que peu d'alternatives satisfaisantes en termes de régulation, car on se trouve souvent piégé entre Charybde et Scylla. La régulation des hypertrucages et autres technologies de l'IA utilisés à des fins politiques, par exemple, soulève la question de savoir où se situe la frontière entre les manipulations problématiques et non problématiques. Le fait de simuler une foule massive lors d'un rassemblement est-il répréhensible au point de justifier une interdiction ? Est-ce que retoucher les traits de son visage pour paraître plus abordable est très différent de corriger ses micro-expressions sur Photoshop ? Même en ce qui concerne les hypertrucages utilisés contre un candidat politique, on ne sait pas toujours où se situe la délimitation entre une vidéo satirique anodine et une ingérence dans la campagne électorale. De même, ce qui doit être considéré comme « faux » et ce qui est « réel » constitue souvent un point de divergence entre les électeurs de différents partis politiques. Interdire les « fausses » informations peut facilement devenir un instrument de pouvoir entre les mains des dirigeants en place, leur permettant d'entraver abusivement la liberté d'expression ; mais inversement, ne rien faire peut donner lieu à des ingérences dans le déroulement des processus électoraux. Un grand nombre de questions similaires se pose également pour les nouveaux médias.



Au regard de la multitude d'outils, de produits et de services basés sur l'IA qui se sont démocratisés, tandis que d'autres intègrent d'ores et déjà l'IA dans leur conception fonctionnelle, on voit difficilement comment éviter l'essor des contenus synthétiques et, partant, l'avènement d'une ère post-vérité. Bien qu'il existe plusieurs façons de traiter cette question, chacune comporte sa propre problématique. Prenons, par exemple, l'hypothèse selon laquelle l'IA inciterait les fournisseurs d'hébergement, les plateformes et les services de contenus à exécuter des programmes de détection de l'IA pour filtrer les contenus générés et manipulés par l'IA. Or, cela reviendrait à bloquer également une multitude de contenus légitimes et anodins, puisque la plupart des contenus seront (le plus souvent de façon marginale) manipulés par l'IA et que même les contenus substantiellement modifiés par l'IA et générés par l'IA sont souvent légitimes. Une autre option consisterait à se fier uniquement aux contenus comportant un filigrane, tout au moins lorsqu'ils sont utilisés par les médias ou dans le cadre de procédures judiciaires ; un filigrane est une notice intégrée à une photo, une vidéo ou tout autre document pour signaler ce qui a été modifié ou manipulé, quand et par qui. Toutefois, du fait de l'exigence d'un filigrane, certains éléments de preuve potentiellement authentiques susceptibles de disculper une personne pourraient être déclarés non recevables par un tribunal, et des documents importants présentant un intérêt majeur pourraient être négligés par les médias.

Quant à la disparition de la confrontation, elle ne porte pas atteinte à l'autonomie des personnes, au contraire. Elle s'appuie sur la réaction profondément instinctive des êtres humains visant à éviter toute expérience pénible ou insoutenable. Un régulateur qui imposerait de force l'expérience de la confrontation à la population au motif que c'est pour son bien court le risque d'être excessivement paternaliste. Il est possible de fixer des limites aux technologies dans certains contextes spécifiques, en interdisant, par exemple, aux étudiants d'utiliser ChatGPT pour écrire une dissertation, mais même dans ce cas, cela pose un certain nombre de questions complexes, telles que, notamment : est-il possible de faire respecter une telle interdiction ? Les étudiants peuvent-ils demander à ChatGPT de trouver des informations pertinentes ou des ouvrages de référence ? *Quid* des locuteurs qui parlent dans une langue autre que leur langue maternelle et souhaitent remédier à leur handicap linguistique en ayant recours à un LLM ? Par ailleurs, il n'y a pas de démarcation claire entre les cas où le remplacement des humains par l'IA est utile, par exemple lorsque les interprètes d'un film n'ont pas à effectuer de cascades dangereuses, et les nombreux cas qui justifient la crainte de voir un remplacement massif des humains par les machines dans le monde du travail. De plus, il n'y a souvent pas de différenciation nette en matière d'originalité et de créativité lorsque les humains utilisent l'IA à des fins de co-créativité entre humains et IA.

Enfin, le régime juridique actuel n'est pas conçu pour protéger des entités non humaines. La GenAI soulève diverses questions à cet égard. Tout d'abord, la plupart des régimes juridiques ne protègent pas de façon adéquate les avatars créés par des gens, ce qui les laisse souvent sans protection contre le vol et contre la violence virtuelle ou le viol. Ensuite, l'essor des hypertrucages de personnes décédées a donné lieu à un débat concernant leur protection. Beaucoup de gens, par exemple, ne souhaitent pas qu'après leur mort, leur conjoint ou conjointe communique quotidiennement avec leur image hypertruquée ou leur fasse prononcer un discours à leurs propres funérailles, mais il n'existe aucune protection juridique contre une telle pratique. Or, il n'est pas évident de



savoir où tracer la limite à cet égard. Une personne en couple peut, par exemple, communiquer chaque jour avec une photo ou visionner une vidéo en boucle de sa compagne ou de son compagnon ; la communication quotidienne basée sur l'IA semble être juste un cran au-dessus sur une échelle mobile. Par ailleurs, il convient de savoir quelles sont les barrières morales permettant de gérer les représentations hyperréalistes de personnes fictives. Tout en sachant pertinemment que la non-réglementation de cet aspect pose des questions éthiques complexes et pourrait avoir des conséquences préjudiciables dans la vie réelle, il s'agit d'évaluer si une réglementation ne risque pas de trop étouffer la liberté d'expression et la créativité, ou de s'ingérer dans l'intimité des gens pour sanctionner certains actes immoraux. Enfin, on peut se demander si les entités dirigées par l'IA devraient bénéficier d'une protection morale et/ou juridique. Si tel est le cas, il convient d'établir à quel moment les machines seront suffisamment intelligentes et auront une telle conscience de soi qu'elles devront être considérées comme des AGI et des êtres sensibles, ce pour quoi il n'existe actuellement aucun test unanimement reconnu.

Les contenus porno-synthétiques sont déjà interdits par le droit pénal dans la plupart des systèmes juridiques. Étant donné l'énorme volume d'images et de vidéos porno-synthétiques, les autorités chargées de faire respecter la loi ont du mal à s'attaquer au problème, d'autant plus qu'il n'est pas toujours aisé d'établir qui a réalisé la vidéo et que les services permettant sa diffusion sont souvent basés dans des juridictions étrangères.

L'interdiction des technologies, des applications et des services, tout au moins ceux qui font explicitement de la publicité pour des productions à caractère sexuel⁶⁰⁸, pourrait résoudre ce problème, mais c'est une mesure radicale, car elle interdit également les utilisations légitimes et utiles, entrave l'innovation technologique par le biais d'expérimentations fortuites, et à l'instar de toutes les interdictions, en particulier dans l'environnement en ligne, elle peut aisément être contournée.

⁶⁰⁸ <https://www.vice.com/en/article/deepnude-app-creates-fake-nudes-of-any-woman/>



11. Conclusions

Le temps presse. Nous nous trouvons sans aucun doute à un moment charnière de l'évolution technologique et réglementaire. L'intégration de l'IA dans le secteur audiovisuel offre des opportunités et de nouveaux outils d'assistance pour développer l'innovation, la créativité et l'efficacité. Mais elle s'accompagne également de défis qui pourraient nécessiter une action réglementaire.

Le cadre réglementaire actuel est constitué de législations diverses, dont certaines sont encore à l'état embryonnaire. L'efficacité du cadre législatif pour assurer une intégration bénéfique et durable de l'IA dans le secteur audiovisuel reste à démontrer.

Avec la généralisation de l'utilisation de l'IA par les citoyens, que ce soit à des fins récréatives ou pour la restauration de contenus et l'aide à la création, il est impératif que la liberté d'expression (article 10 de la Convention européenne des droits de l'homme) reste une préoccupation majeure dans le cadre de la réglementation de l'IA. Ce travail a déjà commencé. La création par le Conseil de l'Europe d'un comité d'experts sur les implications de l'IA générative pour la liberté d'expression en avril 2024 est une étape importante⁶⁰⁹. Ce comité est chargé de rédiger une note d'orientation non contraignante sur les implications de l'IA générative pour la liberté d'expression d'ici la fin de l'année 2025. Selon le projet de rapport de la réunion du comité d'avril 2024⁶¹⁰, la note d'orientation devrait être articulée autour des bénéfices et des risques systémiques. Les avantages comprennent l'élargissement de l'accès à l'information à grande échelle, l'adaptation du format de l'information à l'individu, par exemple en simplifiant le langage et en communiquant visuellement, ou en permettant une meilleure compréhension et utilisation de l'information ; l'accroissement de la visibilité de différentes voix, en fournissant une plateforme appropriée également pour les groupes et les individus en situation de vulnérabilité⁶¹¹. Les risques englobent la diffusion de la désinformation, le processus de rupture du lien de la personne avec la société, l'exclusion numérique, la manipulation, la tricherie, les « deepfakes » et les aspects environnementaux des modèles de fondation⁶¹².

⁶⁰⁹<https://www.coe.int/fr/web/freedom-expression/msi-ai-committee-of-experts-on-the-impacts-of-generative-artificial-intelligence-for-freedom-of-expression>

⁶¹⁰ Comité d'experts sur les implications de l'IA générative pour la liberté d'expression (MSI-AI), 1^{ère} réunion des 23-24 avril 2024, [Projet de rapport de réunion](#), MSI-AI(2024)04.

⁶¹¹ [Projet de rapport de réunion](#), point 14.

⁶¹² *Ibid.*, point 15.



En outre, à la suite de l'adoption par le Comité des ministres du Conseil de l'Europe de la Convention-cadre sur l'IA et les droits de l'homme, la démocratie et l'Etat de droit en mai 2024, le comité sur l'IA élabore actuellement la première méthodologie non contraignante pour l'évaluation des risques et de l'impact des systèmes d'IA du point de vue des droits humains, de la démocratie et de l'Etat de droit (HUDERIA) afin de soutenir la mise en œuvre de la Convention-cadre⁶¹⁵.

Ces initiatives témoignent de l'importance de la liberté d'expression et la nécessité de trouver un équilibre entre les intérêts des citoyens et l'évolution rapide des technologies de l'IA.

⁶¹⁵ [Mandat du Comité sur l'IA, 1^{er} janvier 2024 – 31 décembre 2025](#)

Une publication
de l'Observatoire européen de l'audiovisuel

