



Droit d'auteur et entraînement de l'IA

IRIS

Une publication
de l'Observatoire européen de l'audiovisuel



IRIS-2

Droit d'auteur et entraînement de l'IA

Observatoire européen de l'audiovisuel, Strasbourg, 2026

ISSN 2079-1062

Directrice de publication – Pauline Durand-Vialle, Directrice exécutive

Supervision éditoriale – Maja Cappello, Responsable du Département Informations juridiques

Observatoire européen de l'audiovisuel

Équipe éditoriale – Maja Cappello, Sophie Valais, Diego de la Vega

Observatoire européen de l'audiovisuel

Auteur

Diego de la Vega

Traducteur

Marco Polo Sarl

Assistante éditoriale – Alexandra Ross

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

Observatoire européen de l'audiovisuel

Éditeur

Observatoire européen de l'audiovisuel

76, allée de la Robertsau, 67000 Strasbourg, France

Tel.: +33 (0)3 90 21 60 00

iris.obs@coe.int

www.obs.coe.int

Maquette de couverture – ALTRAN, France

Veuillez citer cette publication comme suit :

D. De la Vega, *Droit d'auteur et entraînement de l'IA*, IRIS, Observatoire européen de l'audiovisuel, Strasbourg, juillet 2026

© Observatoire européen de l'audiovisuel (Conseil de l'Europe), Strasbourg, 2026

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.

Afin de favoriser un langage vecteur d'inclusivité, nous suivons les [lignes directrices du Conseil de l'Europe](#) et privilégions, dans la mesure du possible, l'emploi des mots et expressions épicènes.

Droit d'auteur et entraînement de l'IA

Diego de la Vega



Avant-propos

Plus de trois siècles se sont écoulés depuis que la loi de la Reine Anne (*Statute of Anne*) a jeté les bases du droit d'auteur moderne en s'efforçant de parvenir à un juste équilibre entre les droits des auteurs et ceux des éditeurs. Son intitulé complet précisait à la fois son objectif de texte de loi pour « *l'encouragement de l'apprentissage* » et les moyens mis en œuvre pour atteindre cet ambitieux objectif « *en dévoluant les copies de livres imprimés aux auteurs ou aux acheteurs de ces exemplaires, pendant les périodes qui y sont mentionnées* ». Le premier paragraphe décrivait ensuite un problème présenté comme nouveau mais qui, en réalité, est aussi vieux que le monde : « [...] *Des imprimeurs, des libraires et d'autres personnes se sont récemment et fréquemment permis d'imprimer, de réimprimer et de publier [...] des livres et autres écrits, sans le consentement des auteurs ou des propriétaires [...], à leur très grand préjudice, et bien trop souvent jusqu'à leur ruine et celle de leurs familles* ».

La loi de la Reine Anne de 1710 portait principalement sur les livres imprimés et les éditeurs peu scrupuleux, tandis que la loi du 13 janvier 1791 relative aux spectacles, promue par Pierre-Augustin Caron de Beaumarchais et d'autres dramaturges français, traitait de la représentation publique non autorisée de pièces de théâtre. Chaque nouvelle époque a ensuite apporté ses propres avancées technologiques, ses problèmes et, finalement, ses propres solutions pratiques et juridiques.

Nous voici désormais à l'ère de l'intelligence artificielle (IA), une avancée technologique révolutionnaire qui confronte le droit d'auteur à une multitude de problèmes complexes, qui font ironiquement écho au texte de la loi de la Reine Anne, ou du moins à un mot particulier de son intitulé :

L'apprentissage.

Bien entendu, la loi de la Reine Anne visait à encourager *l'apprentissage* et se préoccupait tout particulièrement des profiteurs sans scrupules qui exploitaient *les livres et autres écrits sans autorisation*. Grâce à ce texte, un nouveau droit (d'auteur) a encouragé *l'apprentissage* en garantissant un moyen de subsistance aux auteurs ou aux titulaires de droits qui, autrement, auraient été spoliés.

Toutefois, avec l'avènement de l'intelligence artificielle, les mots en italique mentionnés ci-dessus ont, d'une certaine manière, pris un nouveau sens. *L'apprentissage* est désormais le fait de machines, et les personnes qui les contrôlent insistent pour que cet *apprentissage* se fasse *sans le consentement des auteurs ou des titulaires de droits*, et probablement sans se soucier particulièrement du *préjudice considérable occasionné à ces derniers*.

Apprentissage. Consentement. Préjudice.

Le présent rapport aborde ces trois concepts en examinant les questions juridiques qui se posent lors de l'entraînement des systèmes d'IA. À cette fin, il explore la manière dont les grands principes du droit d'auteur peuvent s'appliquer aux phases d'alimentation de l'IA, ainsi que l'actuelle interaction entre le droit d'auteur et l'IA générative et agentique. Il vise ainsi à fournir une base équilibrée et documentée pour comprendre l'un des thèmes déterminants de la politique en matière de droit d'auteur, non seulement en Europe, mais

également dans le monde entier. Une première observation, qui peut paraître évidente, consiste à rappeler que les approches en matière de droit d'auteur et d'IA restent très différentes selon que l'on se situe d'un côté ou de l'autre de l'Atlantique.

Bonne lecture !

Maja Cappello

Coordinatrice IRIS

Responsable du Département informations juridiques

Observatoire européen de l'audiovisuel

Table des matières

1. La production de contenus par l'IA dans son contexte et la question du droit d'auteur.....	1
1.1. L'intelligence artificielle et la production de contenus en Europe : état des lieux	1
1.2. Les fonctionnalités de l'intelligence artificielle et le processus de génération de contenus	4
1.3. Une stratégie technologique de création de contenus par l'IA.....	6
1.4. L'interaction entre le droit d'auteur et l'intelligence artificielle	9

2. Le rôle des exceptions en matière de fouille de textes et de données dans l'entraînement de l'IA.....	12
2.1. La réglementation de l'IA : le cadre général européen applicable à l'IA	12
2.2. Le cadre applicable au droit d'auteur.....	13
2.2.1. Le cadre européen du droit d'auteur et la réglementation applicable à l'entraînement de l'IA.....	14
2.2.2. La fouille de textes et de données et les obligations en matière de transparence.....	19
2.2.3. Sélection d'exemples de stratégies en matière de protection du droit d'auteur dans le domaine de l'IA.....	21

3. La question du droit d'auteur dans l'entraînement de l'IA et dans les instructions données à l'IA.....	26
3.1. Les activités d'entraînement de l'IA.....	26
3.2. Les instructions (prompts) destinées à l'IA.....	28

4. Plateformes et utilisateurs : les conditions générales d'utilisation	30
--	-----------

5. Les principaux enseignements	33
--	-----------

Illustrations

Figure 1. Le processus de l'IA agentique 7

Figure 2. Synthèse des technologies de base utilisées par l'IA et l'IA générative..... 9

Figure 3. Le processus technique d'entraînement des systèmes d'IA. 28

Tableaux

Table 1. Les droits d'auteur concernés par l'entraînement des systèmes d'IA 15

Table 2. Les exceptions pertinentes en matière de fouille de textes et de données dans le processus d'entraînement de l'IA..... 18

Table 3. Entraînement de l'IA. Les différentes stratégies possibles..... 23

Table 4. Les stratégies européennes et nationales d'entrainement de l'IA..... 24

Table 5. Dispositions relatives à l'entraînement de l'IA dans les conditions d'utilisation de certaines des plateformes d'IA les plus pertinentes. 31



1. La production de contenus par l'IA dans son contexte et la question du droit d'auteur

1.1. L'intelligence artificielle et la production de contenus en Europe : état des lieux

Depuis le lancement de la version publique de ChatGPT par OpenAI le 30 novembre 2022¹, l'utilisation des outils d'intelligence artificielle (IA) s'est considérablement accélérée, que ce soit à des fins privées ou professionnelles. En janvier 2023, deux mois seulement après son lancement, le nombre d'utilisateurs de la plateforme dépassait déjà les 100 millions². De nombreuses autres plateformes d'IA ont ensuite vu le jour, permettant aux utilisateurs de bénéficier d'une multitude de fonctionnalités.

Le fonctionnement de ces plateformes, notamment lorsqu'il s'agit de systèmes d'intelligence artificielle générative (GenAI) et de technologies similaires telles que l'IA agentique (*Agentic AI*), peut être scindé en deux principales étapes. La première concerne l'entraînement et l'utilisation effective des systèmes d'IA, tandis que la seconde consiste à obtenir le résultat souhaité, qui résout un problème concret à travers un résultat final.

Afin d'analyser les conséquences de ces technologies sur le droit d'auteur, le présent rapport se penchera sur la phase de saisie dans les processus génératifs, comme l'entraînement des systèmes d'IA et la saisie d'instructions par les utilisateurs pour obtenir un résultat spécifique. Bien que l'entraînement de l'IA concerne principalement les fournisseurs de systèmes d'IA et que la saisie d'instructions soit principalement du ressort des utilisateurs, ces deux aspects ont, dans l'actuel cadre européen, des implications significatives en matière de droit d'auteur.

Dans ce contexte, les outils d'IA peuvent relever de différentes catégories d'IA en fonction des résultats souhaités par les utilisateurs. Par exemple, l'IA générative a récemment dominé l'utilisation de l'IA, mais des études montrent que le marché s'oriente rapidement vers d'autres innovations telles que l'IA agentique qui est en mesure d'effectuer

¹ Voir le *OpenAI* site web, [Nouveau : ChatGPT](#).

² K. Hu, [ChatGPT sets record for fastest-growing user base - analyst note](#), *Reuters*, 2 février 2023.



des tâches de manière autonome à la place des utilisateurs. En revanche, l'IA générative en tant que telle fait actuellement face à un certain désenchantement passager³.

Toutefois, seul le domaine de l'IA générative dispose aujourd'hui d'une gamme particulièrement étendue d'outils sur un marché bien développé en Europe, où ces services sont généralement proposés par de grandes entreprises technologiques⁴. Près de 20 % des entreprises européennes et 55 % des grandes entreprises⁵ utilisaient déjà l'IA en 2025 (principalement l'IA générative), ce qui représente une augmentation de 6,5 % par rapport à 2024. Il convient également de noter qu'environ 10 % des entreprises européennes, notamment les sociétés de production, ont déjà recours à l'IA pour la production d'images, de vidéos et de fichiers audio⁶. Des différences semblent toutefois commencer à apparaître entre les différents territoires. En termes de répartition géographique, les statistiques montrent que les habitants d'Europe du Nord utilisent davantage l'IA, tandis que les pays du Sud sont à la traîne. Ainsi, 56,3 % des Suédois ont utilisé l'IA générative en 2026, alors que ce chiffre n'atteint que 19,9 % chez les Italiens⁷.

Les événements tels que le lancement de ChatGPT en 2022 ne constituent toutefois qu'une étape importante parmi d'autres dans le développement de l'IA. Une étape certes majeure, mais qui s'inscrit dans un cheminement qui avait été amorcé dès les années 1950⁸. L'évolution constante de l'IA se traduit par l'apparition continue de nouvelles fonctionnalités offertes aux utilisateurs par les systèmes d'IA, qui sont le signe d'un marché particulièrement complexe. Cette technologie, dont les progrès se mesurent en « étés » et en « hivers »⁹, traverse selon les auteurs une période particulièrement faste, communément

³ H. Khandabattu, [Latest hype cycle for artificial intelligence goes beyond GenAI](#), Gartner Foundation, 8 juillet 2025. Selon les différentes étapes du cycle hyperbolique de Gartner, applicables à toutes les technologies émergentes, la GenAI se trouve actuellement dans la phase du « désenchantement » (*Trough of Disillusionment*), où « l'intérêt s'estompe à mesure que les expériences et les mises en œuvre ne donnent pas les résultats escomptés. Les fabricants de cette technologie disparaissent ou font faillite. Les investissements se poursuivent uniquement si les fournisseurs qui réussissent à surmonter cette épreuve améliorent leurs produits afin de satisfaire les premiers utilisateurs ».

⁴ Conférence des Nations unies sur le commerce et le développement (CNUCED), [Technology and Innovation Report 2025](#) [Rapport sur la technologie et l'innovation 2025], (uniquement en anglais), UN-iLibrary, Genève, 2025.

⁵ En Europe, les grandes entreprises sont celles qui emploient 250 personnes ou plus. Eurostat, [Statistiques expliquées](#) (e-traduction).

⁶ Eurostat, [Use of artificial intelligence in enterprises](#) [Utilisation de l'intelligence artificielle dans les entreprises] (en anglais), décembre 2025.

⁷ Eurostat, [The use of artificial intelligence \(AI\) technologies in the European Union](#) [L'utilisation des technologies d'intelligence artificielle (IA) dans l'Union européenne] (en anglais), Key Results, 2026.

⁸ Les fondements de l'IA se trouvent dans l'ouvrage d'Alan Turing intitulé « *Computing machinery and intelligence* », qui visait à déterminer si les machines sont capables de penser ou non. Turing, figure centrale de la technologie numérique, a également mis au point ce qu'on appelle le test de Turing, qui consistait à évaluer la capacité des machines à répondre à des questions posées par des êtres humains, anticipant ainsi le modèle de *chatbot* utilisé aujourd'hui par de nombreuses plateformes d'IA populaires. Le terme « intelligence artificielle » est apparu dès 1956 et des expériences décisives ont déjà été menées tout au long des années 1960. Voir D. Blagoj, T. Chrisa et K. Uroš, « [Historical Evolution of Artificial Intelligence](#) », EUR 30221 EN, Office des publications de l'Union européenne, Luxembourg, 2020, ISBN 978-92-76-18940-4, doi:10.2760/801580, JRC120469.

⁹ E. Francesconi, [The winter, the summer and the summer dream of artificial intelligence in law](#), allocution prononcée lors de la 18e Conférence internationale sur l'intelligence artificielle et le droit, Artificial Intelligence and Law, Springer Nature, 3 février 2022.



désignée comme « l'été de l'IA ». Cette évolution s'explique par l'intérêt croissant, bien que parfois fluctuant, des entreprises et des utilisateurs pour l'évolution de l'IA, contrairement au contexte passé où l'intérêt pour l'IA, dans le cadre d'un développement technologique plus général, était secondaire¹⁰.

Parallèlement, les législateurs du monde entier s'efforcent de définir un cadre réglementaire approprié en abordant les diverses implications juridiques de l'IA. Par exemple, la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle¹¹, premier traité international juridiquement contraignant en matière d'IA, ainsi que des textes législatifs européens pertinents tels que le règlement sur l'intelligence artificielle (*Artificial Intelligence Act* - AIA)¹², ont été adoptés ces dernières années. Parallèlement, depuis juin 2026, la Commission européenne (CE) recueille des données afin d'évaluer l'efficacité du cadre juridique existant et de déterminer si des mesures complémentaires sont nécessaires¹³. Plus précisément, selon l'appel à contributions sur le droit d'auteur, publié en mai 2026, « les travaux préparatoires à l'initiative législative ciblée se dérouleront parallèlement à la révision de la CDSMD. C'est pourquoi un appel à témoignages unique est lancé : il vise à recueillir les informations nécessaires pour étayer la révision de la directive CDSMD, et à recueillir des retours d'expérience sur les défis liés notamment à l'application et au respect du droit d'auteur et des droits voisins dans le contexte des évolutions technologiques, ainsi que sur les moyens potentiels d'y remédier ».¹⁴

Cependant, bien que la législation générale soit déjà en vigueur, la position de la Cour de justice de l'Union européenne sur l'IA générative et le droit d'auteur, qui vise à compléter et à interpréter la législation européenne déjà applicable, doit encore être précisée. En matière de droit d'auteur, son positionnement est particulièrement important aussi puisqu'il n'existe pas de législation spécifique à l'IA générative. Les dispositions relatives au droit d'auteur restent en effet soumises au cadre général du droit d'auteur.

À cet égard, la décision rendue par la Cour de justice de l'Union européenne dans l'affaire *Like Company c. Google Ireland Limited*¹⁵ devrait constituer le tout premier arrêt en

¹⁰ R. Susskind, *The summer of AI* in "How to think about AI: a guide for the perplexed", Oxford Academic, 20 mars 2025. L'auteur affirme que des événements tels que le développement du World Wide Web ont entraîné une forme de stagnation dans le développement de l'IA.

¹¹ Conseil de l'Europe, *Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle* a été ouverte à la signature le 5 septembre 2024 et vise à combler les éventuelles lacunes juridiques susceptibles de découler des rapides évolutions technologiques, sous l'angle des droits de l'homme, de la démocratie et de l'État de droit. Elle est neutre sur le plan technologique et ne régit pas la technologie en soi, afin de ne pas être conditionnée par les évolutions constantes de l'IA.

¹² *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).

¹³ *Directive (EU) 2019/790* du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE.

¹⁴ Commission Européenne, *Appel à contributions en vue d'une analyse d'impact*, Section A, 13 Mai 2026.

¹⁵ *Affaire C-250/25, Like Company*. Demande de décision préjudicielle, 3 avril 2025 – *Like Company c. Google Ireland Limited*. La demande de décision préjudicielle, présentée par le *Környéki Törvényszék* de Budapest (Hongrie), vise notamment à obtenir des éclaircissements sur des questions telles que la qualification des réponses fournies par des *chatbots* basés sur des grands modèles de langage (LLM) en tant que communication



la matière au niveau de l'Union européenne. En relation avec la protection des publications de presse en ce qui concerne les utilisations en ligne¹⁶, elle devrait notamment permettre de préciser certains aspects essentiels de la réglementation européenne applicable en matière de droit d'auteur dans le contexte de l'IA.

Même si aucune décision de principe n'a encore été rendue pour harmoniser les règles existantes, d'autres décisions, comme la récente décision rendue dans l'affaire *Meta Platforms Ireland*¹⁷, peuvent indirectement apporter des précisions sur des aspects essentiels du droit d'auteur et de l'IA – comme le droit exclusif au titre du droit d'auteur prévu par la directive sur le droit d'auteur dans le marché unique numérique, lequel ne saurait se substituer au droit à une rémunération des éditeurs de presse. Celui-ci pourrait également apporter des éléments d'orientation aux différentes juridictions nationales qui ont déjà été amenées à statuer dans des affaires portant sur l'IA et le droit d'auteur.

1.2. Les fonctionnalités de l'intelligence artificielle et le processus de génération de contenus

Comme le reconnaît le règlement sur l'intelligence artificielle (*Artificial Intelligence Act* – AIA), l'IA désigne un ensemble de technologies en pleine évolution. Ce règlement précise que ces technologies peuvent, à terme, offrir un large éventail d'avantages économiques, environnementaux et sociétaux pour l'ensemble du secteur économique et des activités sociales. Leur fonctionnement repose sur l'analyse de grands ensembles de données qui sont susceptibles d'être protégés par le droit d'auteur.

Plus précisément, ces technologies peuvent contribuer à la création de contenus, à l'amélioration des processus de prédiction et à l'optimisation de toutes sortes d'opérations de la vie quotidienne et de l'allocation des ressources ; elles permettent également de personnaliser les solutions numériques mises à la disposition des particuliers et des

au public, la qualification du processus d'entraînement d'un *chatbot* basé sur un grand modèle de langage en tant qu'acte de reproduction, le rôle global des exceptions relatives à la fouille de textes et de données dans le domaine de l'IA, ainsi que certaines techniques de « *prompting* ».

¹⁶ Art. 15 CDSMD.

¹⁷ Cour de justice de l'Union européenne, *Meta Platforms Ireland Ltd c. Autorità per le Garanzie nelle Comunicazioni* [GC], Affaire C-797/23, Cour de justice de l'Union européenne, 12 mai 2026. Au paragraphe 62 de son arrêt, la grande chambre précise que « les États membres ne sauraient transposer l'article 15 de la directive 2019/790 en substituant aux droits exclusifs de nature préventive qu'il établit un simple droit à compensation, qui permettrait aux éditeurs de publications de presse d'obtenir uniquement une rémunération pour les utilisations en ligne de ces publications par des fournisseurs de services de la société de l'information, mais non pas d'interdire ces utilisations ».



organisations¹⁸. Ces fonctionnalités sont proposées soit de manière autonome, soit en tant que composante d'un produit numérique précis¹⁹.

L'intelligence artificielle générative propose plusieurs fonctionnalités, qui varient en fonction du type de problème à résoudre. Les assistants virtuels basés sur la technologie GenAI sont l'une des fonctionnalités les plus répandues. Ils utilisent de grands modèles de langage (*Large Language Models* – LLM) afin de générer des résultats de haute qualité. À ces assistants virtuels s'ajoutent d'autres modèles de systèmes d'IA, comme l'IA générative multimodale, qui est capable de traiter un ensemble d'entrées différentes (images, vidéos ou fichiers audio) pour générer des résultats complexes, comme des contenus de type « texte vers image » ou « script vers vidéo »²⁰.

L'IA agentique, plus récente, peut également être classée comme une forme d'IA générative, dans laquelle les systèmes ne se contentent pas de générer des contenus, mais prennent également des décisions autonomes et mènent des actions spécifiques en lien avec celles-ci sans avoir besoin d'une entrée pleinement explicite. L'IA agentique est capable de réaliser des objectifs complexes, ainsi que de concevoir et d'exécuter des flux de travail avec divers degrés d'intervention humaine. Au cours des trois à six prochaines années, l'IA agentique devrait connaître une phase de forte croissance.

Enfin, parmi le vaste ensemble des technologies d'IA générative en cours de développement, les assistants de codes d'IA, capables de créer des codes logiciels, notamment pour la conception d'applications ou de jeux vidéo, avec un haut degré d'autonomie et des instructions simples, devraient également se généraliser à un rythme particulièrement soutenu²¹.

Chaque secteur économique est susceptible d'être affecté par l'IA à des degrés divers. Pour certains secteurs, l'impact et l'importance du droit d'auteur peuvent s'avérer déterminants, tandis que dans d'autres, ces effets restent encore limités²². Lorsqu'elle peut jouer un rôle majeur, comme dans le secteur audiovisuel²³, notamment pour les tâches

¹⁸ Considérant 6 du règlement sur l'IA. Ce contenu peut également inclure, par exemple, les *deep fakes*, en lien avec les droits de la personnalité et les droits à l'image, définis comme des contenus visuels, audio ou vidéo générés ou manipulés par l'IA qui ressemblent à des personnes, objets, lieux, entités ou événements existants et qui pourraient être perçus comme authentiques ou véridiques par une personne. D'autres implications peuvent concerner, par exemple, la liberté d'expression. Voir la [note d'orientation du Conseil de l'Europe sur les implications de l'IA générative pour la liberté d'expression](#).

¹⁹ Par exemple, les [systèmes d'IA générative et d'IA agentique d'Adobe](#) sont « intégrés dans l'ensemble des applications créatives et commerciales d'Adobe », Copilot de Microsoft est intégré à l'« écosystème Microsoft 365 », tandis que certains des outils d'IA les plus courants sont des *chatbots* autonomes tels que [ChatGPT](#) ou [Claude](#).

²⁰ Zimmermann A. et al., [Emerging Tech Impact Radar: Generative AI](#), Gartner, 14 février 2025.

²¹ *Ibid.*

²² Eurostat, [The use of artificial intelligence \(AI\) technologies in the European Union](#) [L'utilisation des technologies d'intelligence artificielle (IA) dans l'Union européenne], (en anglais), *Key Results*, 2026. Par exemple, certains secteurs, comme celui de la construction, ne sont pas autant touchés par l'IA que d'autres, notamment les industries de la création, où l'IA, et en particulier l'IA générative, peut prendre en charge un grand nombre de tâches.

²³ Pour un aperçu spécifique de l'IA dans le secteur audiovisuel, voir M. Cappello (sous la direction de), « [L'IA et le secteur audiovisuel : exploration du paysage juridique actuel](#) », IRIS, Observatoire européen de l'audiovisuel, Strasbourg, octobre 2024.



purement créatives²⁴, ou dans le développement de jeux vidéo²⁵, l'IA semble déjà avoir des répercussions considérables. En règle générale, les activités à forte composante cognitive sont davantage susceptibles d'être directement concernées par l'IA. Ces secteurs devraient ainsi voir les effets de l'IA, déjà présents dans des activités non liées à l'IA générative, s'amplifier grâce à l'IA générative et à la création de contenus de synthèse²⁶.

1.3. Une stratégie technologique de création de contenus par l'IA

L'IA générative est une forme particulière d'intelligence artificielle, qui se caractérise par sa capacité à accomplir des tâches variées. D'un point de vue technique, elle utilise des requêtes (*prompts*) et le langage humain naturel, et repose sur l'apprentissage profond (*Deep Learning* – DL), à savoir une sous-catégorie de l'apprentissage automatique (*Machine Learning* – ML), qui permet de réaliser des tâches complexes grâce à l'utilisation de vastes ensembles de données et à un entraînement adapté, avec pour objectif la création de nouveaux contenus²⁷.

D'un point de vue juridique, le Parlement européen a récemment adopté une résolution sur le droit d'auteur et l'intelligence artificielle générative (également connue sous le nom de « rapport Voss »), dans laquelle il définit l'IA générative comme « un type d'IA qui, contrairement à d'autres systèmes d'IA conçus principalement pour classer ou prédire, génère de nouveaux contenus, tels que du texte, des images, de la musique, des vidéos et du code, sur la base d'un entraînement utilisant de très grands ensembles de données à partir desquels ils apprennent des systèmes et des structures ». Pour ce faire, les systèmes d'IA générative utilisent des prédictions basées sur des modèles statistiques, qui sont capables d'imiter la créativité humaine et reposent sur des contenus préexistants²⁸.

De manière plus générale, le règlement sur l'IA décrit simplement les modèles d'IA à usage général comme une technologie permettant la production flexible de contenus –

²⁴ À titre d'exemple, voir les lignes directrices de Netflix intitulées « [Using Generative AI in Content Production](#) » (Utilisation de l'IA générative dans la production de contenus) destinées aux entreprises partenaires qui produisent des contenus audiovisuels pour Netflix. En matière de prompts, ces lignes directrices encouragent à éviter d'utiliser des données d'entrée (par exemple, des *prompts* ou des images) faisant référence à des œuvres protégées par le droit d'auteur ou à l'image de personnalités publiques ou de personnes décédées sans les autorisations appropriées.

²⁵ Heinisch O. et Jadot C., « [Some Implications of the EU AI Act on Video Game Developers](#) », Sheppard, 28 février 2025.

²⁶ OCDE, [Macroeconomic productivity gains from Artificial Intelligence in G7 economies](#) [Gains de productivité macroéconomiques liés à l'intelligence artificielle dans les économies du G7], (en anglais), Documents de l'OCDE sur l'intelligence artificielle, juin 2025, n° 41.

²⁷ Voir Microsoft, [IA générative c. autres types d'IA](#), (en anglais).

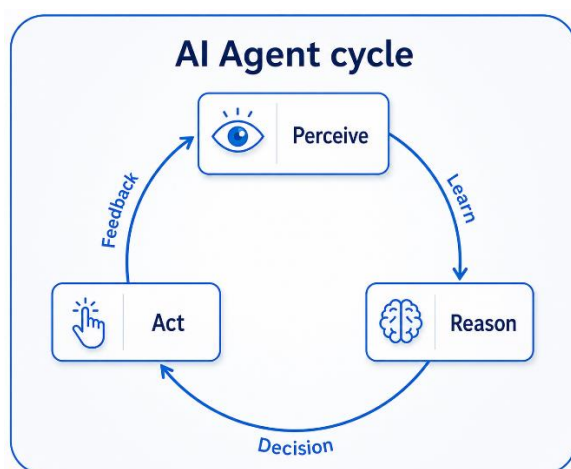
²⁸ Parlement Européen, [Le droit d'auteur et l'intelligence artificielle générative - opportunités et défis](#), Résolution du 10 mars 2026 sur le droit d'auteur et l'intelligence artificielle générative – opportunités et défis (2025/2058(INI)) (Rapport Voss).

sous forme de texte, d'audio, d'images ou de vidéo –, qui sont aisément adaptables à un large éventail de tâches spécifiques²⁹.

D'autres classifications des technologies d'IA, telles que l'IA agentique, restent pour l'instant des questions d'ordre commercial et technique, qui ne font l'objet d'aucune démarche juridique³⁰ quant à leur définition en Europe. Dans le secteur, les IA agentiques sont définies comme des systèmes d'IA capables d'englober un large éventail de fonctions autres que le traitement du langage naturel, parmi lesquelles la prise de décision, la résolution de problèmes, l'interaction avec des environnements externes et l'exécution d'actions³¹.

En dehors de l'Europe, l'Australie vient également de définir l'IA agentique comme un « système basé sur un logiciel qui perçoit les données de son environnement, maintient ou actualise son fonctionnement interne, et sélectionne et exécute des actions à l'aide d'API ou d'autres outils pour atteindre des objectifs définis dans le cadre des autorisations et des contraintes qui lui sont attribuées. Il est capable de planifier, de tirer des enseignements des retours d'information lorsqu'il y est autorisé, et contrôler lui-même son fonctionnement afin d'améliorer ses performances au fil du temps, tout en respectant des mécanismes clairement définis de gouvernance, de surveillance et de contrôle ». ³² L'IA agentique peut également, à terme, générer des contenus.

Figure 1. Le processus de l'IA agentique



Source : Gouvernement australien

²⁹ Considérant 99 du règlement sur l'IA.

³⁰ D'un point de vue informatif, le Contrôleur européen de la protection des données définit l'IA agentique comme « un concept relevant de l'intelligence artificielle (IA) qui décrit des systèmes agissant de manière autonome, avec des interactions humaines limitées (notamment sans instructions détaillées), afin d'atteindre des objectifs plutôt que d'accomplir des tâches isolées ». Goldstein A., [Agentic AI](#) (en anglais), European Data Protection Supervisor

³¹ Gutowska A., [Qu'est-ce qu'un agent IA ?](#) (en anglais), Think, IBM.

³² Gouvernement australien (Digital Transformation Agency), [Agentic AI addendum to the AI technical standard for Australian Government](#), [digital.gov.au](#), dernière mise à jour le 4 juin 2026



L'apprentissage automatique repose, d'une part, sur l'apprentissage d'ensembles de données préalablement existants à l'aide d'algorithmes et, d'autre part, sur l'inférence de nouvelles données. Ce processus comprend notamment l'apprentissage profond qui, selon les spécialistes du secteur, est une « série complexe d'équations imbriquées qui associe une entrée à une sortie »³³. Enfin, les modèles de base sont des modèles d'apprentissage profond spécifiques, entraînés sur de grandes quantités de données disponibles³⁴, qui permettent aux grands modèles de langage d'interagir avec les personnes sous forme de conversations³⁵.

La législation qui régit l'IA, que ce soit du point de vue du droit d'auteur ou dans une perspective plus large, présuppose que les différents niveaux techniques qui composent l'IA reposent nécessairement sur les principes décrits ci-dessus. Toutefois, les nouvelles technologies qui émergent, telles que l'IA quantique³⁶, pourraient bien à l'avenir remettre en question ces principes, puisque la manière dont les systèmes d'IA sont entraînés pourrait s'avérer radicalement différente si le traitement quantique et l'IA venaient à se combiner³⁷.

³³ Bergmann D., [Qu'est-ce que le deep learning ?](#), Think, IBM.

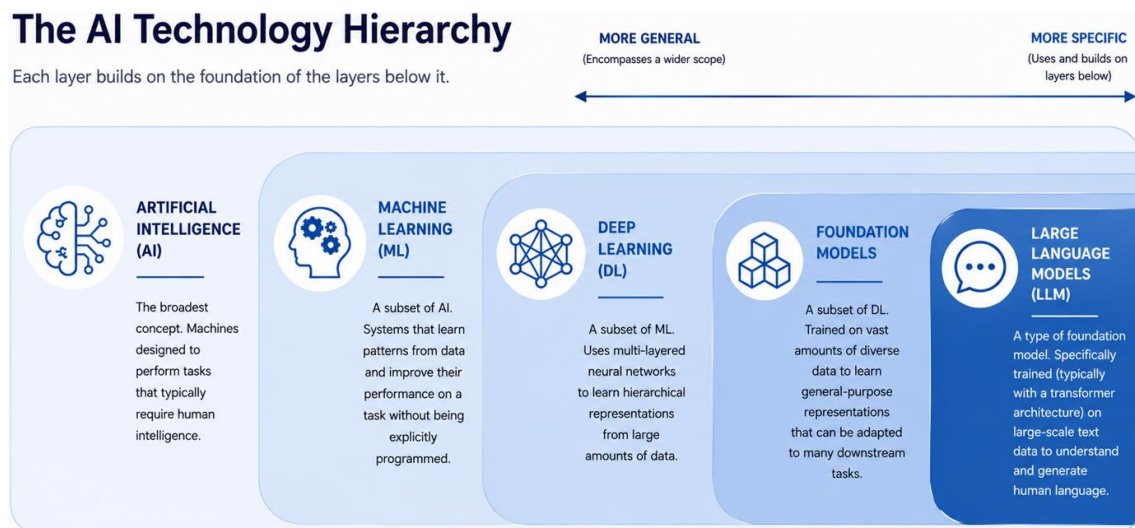
³⁴ Voir [Defining AI and chatbots](#), Stanford Teaching Commons, Université de Stanford. Voir également [Qu'est-ce que le deep learning ?](#), et M. Cappello (sous la direction de), « [L'IA et le secteur audiovisuel : exploration du paysage juridique actuel](#) », *op. cit.*

³⁵ Jurafsky D. et Martin J. H., [Speech and Language Processing](#) (3^e édition – version préliminaire), *Université de Stanford*, chapitre 7.

³⁶ Selon la définition d'Amazon Web Services, l'IA quantique désigne l'utilisation de technologies quantiques pour faire fonctionner des systèmes d'IA, par opposition à l'IA classique, qui repose sur l'informatique ordinaire. Cette technologie pourrait, en théorie, remplacer les normes techniques actuellement en vigueur et améliorer le fonctionnement des systèmes et des plateformes d'IA. Voir le *Amazon Web Services (AWS)* site web, « [Qu'est-ce que l'IA quantique ?](#) ».

³⁷ Acampora G., et autres, [Quantum artificial intelligence: a survey](#), *Computer Science Review* 59, février 2026.

Figure 2. Synthèse des technologies de base utilisées par l'IA et l'IA générative



Source : Observatoire européen de l'audiovisuel

1.4. L'interaction entre le droit d'auteur et l'intelligence artificielle

Puisque l'IA générative est en mesure de créer des contenus, elle soulève d'ores et déjà des questions de droit d'auteur dans le cadre même de son processus de création. L'une des premières préoccupations relatives à l'utilisation de l'IA générative depuis 2022, et à ses nouveaux outils qui ont fait leur apparition sur le marché, concerne notamment l'utilisation de contenus préexistants pour l'entraînement à grande échelle des systèmes d'IA. Par exemple, dès 2023, la grève des scénaristes aux États-Unis a clairement montré que, lorsque des outils d'IA sont utilisés dans l'industrie audiovisuelle, des conflits liés au droit d'auteur entre producteurs et auteurs risquent rapidement de survenir³⁸.

Dans le cadre du Bouclier européen de la démocratie et du secteur des médias, les institutions de l'UE ont conjointement mis en garde contre les répercussions de l'entraînement de l'IA. Tout en saluant le développement de l'IA, les institutions de l'UE avertissent que « l'utilisation de matériel protégé par le droit d'auteur pour former des modèles d'IA sans autorisation demeure une menace importante pour le secteur des médias, car ils érodent les revenus et ont une incidence sur la qualité et la diversité des médias »³⁹.

Comme l'indique le « rapport Voss » du Parlement européen précédemment mentionné, l'entraînement de l'IA « peut inclure du matériel protégé par le droit d'auteur ». Le rapport précise que « le développement, le déploiement et l'utilisation de l'IA doivent

³⁸ Cullins A. et Kilkenny K., [As Writers Strike, AI Could Covertly Cross the Picket Line](#), *The Hollywood Reporter*, 3 mai 2023.

³⁹ Commission Européen, [Communication conjointe au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions. Bouclier européen de la démocratie : renforcer la position de démocraties fortes et résilientes](#), EUR-Lex, JOIN(2025) 791 final, Bruxelles, 12 novembre 2025.



être pleinement conformes au cadre juridique actuel ». Il dresse un tableau relativement pessimiste de la situation mondiale, en particulier pour l'UE, où l'on constate que « des violations généralisées des droits d'auteur par les fournisseurs d'IA générative sont avérées, notamment la collecte non autorisée d'œuvres sur l'internet », ainsi que des violations graves telles que le refus de se conformer aux réserves des titulaires de droits en matière de fouille de textes et de données, l'utilisation de sources piratées pour obtenir des œuvres et l'absence de demande de licence pour l'utilisation d'œuvres préexistantes⁴⁰.

L'Assemblée parlementaire du Conseil de l'Europe (APCE) a récemment fait part de préoccupations similaires. Elle estime en effet que « l'avènement de l'ère de l'intelligence artificielle a fait surgir une série de problèmes particulièrement complexes pour le secteur de la création [dans la mesure où] pour alimenter leurs systèmes très gourmands en données, les entreprises d'IA récupèrent des informations sur internet sans autorisation préalable et sans rémunérer les créateurs de contenus, sur la base de dispositions législatives qui ne sont ni claires ni adaptées à leur objectif »⁴¹.

D'un point de vue juridique, cette situation nécessite effectivement l'application du droit d'auteur. L'entraînement des systèmes d'IA comporte généralement, parmi d'autres actes et techniques, des opérations massives de reproduction et, à terme, l'extraction de contenus à partir de bases de données⁴². Ces opérations sont étroitement liées à la fouille de textes et de données, qui permet le traitement de grandes quantités d'informations en vue d'acquérir de nouvelles connaissances et de découvrir de nouvelles tendances⁴³ afin d'obtenir des informations spécifiques pertinentes pour l'entraînement⁴⁴. Le « *prompting* » pourrait relever d'une situation similaire. Plus précisément, il reste difficile de déterminer si le fait de fournir une entrée sous la forme d'une instruction donnée par l'utilisateur, lorsqu'elle repose sur des contenus protégés par le droit d'auteur intégrés au cours du processus d'entraînement, comme des publications de presse, constitue nécessairement un acte de reproduction de contenus protégés par le droit d'auteur⁴⁵.

L'évaluation de l'impact du droit d'auteur et de l'IA récemment réalisée par le Gouvernement britannique a permis de mettre en évidence le juste équilibre entre l'ensemble des droits et intérêts en jeu. Elle rappelle l'importance de la législation sur le droit d'auteur pour pérenniser le succès des industries du secteur de la création, puisqu'elle « permet aux créateurs d'obtenir une reconnaissance et de tirer un bénéfice financier de leur œuvre, tout en attirant des investissements pour l'avenir ». Toutefois, il est explicitement admis que « les titulaires de droits déclarent rencontrer des difficultés pour contrôler l'utilisation et l'accès à leurs œuvres. Les développeurs d'IA affirment quant à eux

⁴⁰ Considérant L du rapport Voss.

⁴¹ [Résolution 2654 \(2026\)](#), La protection du droit d'auteur dans l'environnement de l'intelligence artificielle, adopté le 23 avril 2026.

⁴² Certaines plateformes en ligne offrent aux utilisateurs la possibilité de retracer en partie l'utilisation de contenus préexistants dans l'entraînement des systèmes d'IA. Par exemple, [AI Watchdog](#), du magazine The Atlantic, propose une « enquête sur les livres, vidéos et autres supports utilisés par les entreprises technologiques les plus puissantes au monde pour entraîner leurs modèles d'IA ».

⁴³ Considérant 8 de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

⁴⁴ Rosati E., [The future of the movie industry in the wake of generative AI: A perspective under EU and UK copyright law](#), *Security Review*, Vol. 59, novembre 2025.

⁴⁵ [Affaire C-250/25](#), *Like Company*, demande de décision préjudicielle, 3 avril 2025 – *Like Company c. Google Ireland Limited*.



avoir du mal à accéder à de grandes quantités de données pour entraîner les modèles de base [...] »⁴⁶.

La réglementation sur le droit d'auteur joue par conséquent un rôle fondamental dans la conciliation des intérêts des titulaires de droits et de ceux des développeurs d'IA ; le secteur des plateformes s'étant quant à lui développé grâce à la fourniture de services d'IA. L'IA générative, l'IA agentique et d'autres technologies assimilées sont au cœur des préoccupations des législateurs, car leurs effets sur l'entraînement des IA et, dans une moindre mesure, sur la génération de *prompts* peuvent avoir un impact considérable non seulement en termes de sécurité juridique, mais également sur le plan économique et sociétal.

⁴⁶ Gov.UK, [Copyright and AI Impact Assessment](#) (Évaluation de l'impact du droit d'auteur et de l'IA). Présentée au Parlement en vertu de l'article 135 de la loi britannique relative aux données (utilisation et accès) de 2025, Ministère des Sciences, de l'Innovation et de la Technologie et ministère de la Culture, des Médias et des Sports, mars 2026 (uniquement en anglais).



2. Le rôle des exceptions en matière de fouille de textes et de données dans l'entraînement de l'IA

2.1. La réglementation de l'IA : le cadre général européen applicable à l'IA

L'Eurobaromètre spécial consacré à la Décennie numérique 2026 révèle que 80 % des Européens sont favorables à une réglementation rigoureuse de l'IA qui puisse garantir la sécurité, plutôt que son évolution sans aucune restriction⁴⁷. Ce sentiment partagé par les Européens s'inscrit dans la logique du règlement sur l'IA, selon lequel « compte tenu de l'incidence majeure que l'IA peut avoir sur nos sociétés et de la nécessité de bâtir la confiance, l'IA et son cadre réglementaire doivent impérativement être élaborés dans le respect des valeurs de l'Union »⁴⁸. Plus généralement, les objectifs 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 visent également à « garantir que les activités menées dans le cadre du cycle de vie des systèmes d'intelligence artificielle sont pleinement compatibles avec les droits de l'homme, la démocratie et l'État de droit »⁴⁹.

Au sein de l'Union européenne, le règlement sur l'IA doit être considéré comme un élément faisant partie d'une stratégie globale qui va au-delà de la réglementation, et qui regroupe des projets d'innovation ainsi qu'une stratégie plus générale au niveau politique⁵⁰.

Le règlement sur l'IA, dont l'application n'est pas encore totalement effective⁵¹, reste aujourd'hui en Europe l'outil réglementaire incontournable en matière d'IA. Il ne traite pas spécifiquement du droit d'auteur, mais se concentre sur le fonctionnement du marché

⁴⁷ Commission européenne : Direction générale de la communication et Direction générale des réseaux de communication, des contenus et des technologies, [Digital decade – Eurobarometer report – February – March 2026](#) (La décennie numérique – Rapport Eurobaromètre – février-mars 2026), uniquement en anglais, Direction générale des réseaux de communication, des contenus et des technologies, 2026. Les résultats font apparaître des disparités territoriales. Alors que les citoyens grecs, luxembourgeois et irlandais sont favorables à une réglementation de l'IA, ceux de Roumanie, de Pologne et d'Estonie se montrent moins enthousiastes, même si une majorité d'entre eux préfère tout de même une réglementation à une démarche plus permissive. Le rapport révèle également que les femmes sont plus enclines que les hommes à se prononcer en faveur d'une réglementation de l'IA et que, dans l'ensemble, les jeunes Européens y sont moins disposés. Voir p. 60 et 61.

⁴⁸ Considérant 6 du règlement sur l'IA.

⁴⁹ 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](#), Article 1.

⁵⁰ Commission européenne, [Approche européenne de l'intelligence artificielle](#), (e-traduction).

⁵¹ Le règlement sur l'IA entrera en vigueur le 2 août 2026, bien que certains chapitres soient déjà applicables depuis 2025 et que l'article 6.1 (relatif aux critères de classification des systèmes d'IA à haut risque) ne s'appliquera qu'à compter du 2 août 2027.



intérieur et la protection de la santé, de la sécurité et des droits fondamentaux⁵². Au niveau national, plusieurs pays disposent de législations sur l'IA déjà en vigueur, comme le Danemark, la Finlande, l'Italie et la Slovaquie, ou ont présenté des propositions législatives en ce sens⁵³.

Toutefois, en matière de droit d'auteur, hormis certaines dispositions spécifiques du règlement sur l'IA en relation avec les modèles d'IA à usage général, comme les obligations de transparence imposées aux fournisseurs de systèmes d'IA, le règlement fait directement référence à la directive sur le droit d'auteur dans le marché unique numérique. Par exemple, les considérants 105 et 106 du règlement sur l'IA mentionnent en effet expressément cette directive, dans la mesure où elle « a introduit des exceptions et des limitations autorisant les reproductions et extractions d'œuvres ou d'autres objets protégés aux fins de la fouille de textes et de données, sous certaines conditions »⁵⁴. Ainsi, le règlement sur l'IA confère à la directive sur le droit d'auteur dans le marché unique numérique d'importantes prérogatives en matière de réglementation de l'IA dans le domaine du droit d'auteur.

2.2. Le cadre applicable au droit d'auteur

La directive sur le droit d'auteur dans le marché unique numérique s'avère être l'un des derniers éléments en date du relativement complexe cadre européen applicable au droit d'auteur⁵⁵. Elle vient compléter plusieurs directives existantes dans lesquelles figurent les principales dispositions relatives aux systèmes nationaux de droit d'auteur, notamment la directive 2001/29/CE relative à l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information (directive « InfoSoc »)⁵⁶. Cette directive a instauré dès 2001 certains des principaux piliers sur lesquels repose encore aujourd'hui le cadre européen du droit d'auteur, et la directive sur le droit d'auteur dans le marché unique numérique en est le prolongement.

La directive sur le droit d'auteur dans le marché unique numérique constitue la pièce maîtresse de la réglementation en matière de droit d'auteur concernant l'IA, puisqu'il n'existe aucune disposition spécifique en la matière ; son raisonnement s'inscrit toutefois dans la continuité de l'ensemble des précédentes directives sur le droit d'auteur⁵⁷.

⁵² Article 1 du règlement sur l'IA.

⁵³ Bashir E., *Status of national AI laws across Europe* [État des lieux des législations nationales en matière d'IA en Europe] (en anglais), *Cullen International*, 12 juin 2026.

⁵⁴ Considérant 105 du règlement sur l'IA.

⁵⁵ Ce cadre général s'applique à tous les secteurs. Par exemple, dans le secteur audiovisuel, il peut concerner la rédaction assistée de scénarios, le montage vidéo assisté par l'IA, le sous-titrage et le doublage, la création de vidéos à la demande, la recommandation de contenus et bien d'autres applications encore.

⁵⁶ *Directive 2001/29/CE* (directive InfoSoc) du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information.

⁵⁷ Outre la directive InfoSoc, les textes réglementaires qui sont susceptibles d'avoir un impact sur l'IA sont notamment les suivants : la directive *96/9/CE* du Parlement européen et du Conseil du 11 mars 1996 concernant la protection juridique des bases de données, la directive *2006/115/CE* relative au droit de location et de prêt



2.2.1. Le cadre européen du droit d'auteur et la réglementation applicable à l'entraînement de l'IA

La directive sur le droit d'auteur dans le marché unique numérique, qui est antérieure à l'émergence de l'IA générative vers 2022-2023⁵⁸, ne fait aucune mention de l'IA mais reconnaît que « les nouvelles technologies permettent une analyse informatique automatisée d'informations sous forme numérique, telles que des textes, des sons, des images ou des données, ce que l'on appelle généralement la fouille de textes et de données », en faisant également référence à « l'élaboration de nouvelles applications ou technologies »⁵⁹. Selon le considérant 3 CDSMD, « La législation en la matière doit résister à l'épreuve du temps afin de ne pas entraver l'évolution des technologies. » Dès l'adoption de la directive CDSMD, le Parlement européen a déclaré que « En outre, outre l'exception obligatoire relative à l'exploration de textes et de données à des fins de recherche proposée par la Commission dans sa proposition, les colégislateurs ont convenu d'inscrire dans le droit de l'Union une autre exception obligatoire concernant l'exploration générale de textes et de données (article 4), afin de contribuer au développement de l'analyse de données et de l'intelligence artificielle, ainsi qu'à l'adoption d'un certain nombre de nouvelles règles de l'Union (par exemple, un nouveau mécanisme d'octroi de licences pour les œuvres qui ne sont plus commercialisées). »⁶⁰

Comme le souligne l'appel à contribution récemment lancé par la Commission européenne pour le réexamen de la directive (UE) 2019/790, « la directive sur le droit d'auteur dans le marché unique numérique et les dispositions relatives au droit d'auteur du règlement sur l'IA et les mesures de mise en œuvre ultérieures, contribuent à soutenir les titulaires de droits dans l'exercice de ces droits. Cependant, ce cadre peut ne pas être suffisant pour remédier aux difficultés rencontrées par les titulaires de droits pour exercer un contrôle et octroyer des licences sur leurs œuvres à des fins d'utilisation par l'IA »⁶¹. Les institutions européennes avaient d'ailleurs déjà annoncé fin 2025 que la Commission réexaminerait la directive sur le droit d'auteur dans le marché unique numérique et étudierait les moyens d'améliorer son efficacité dans le contexte de l'IA générative. En particulier, « Le développement de l'IA générative, tout en offrant des possibilités, est source de plusieurs défis pour l'écosystème des médias. Le piratage en ligne et l'utilisation de matériel protégé par le droit d'auteur pour former des modèles d'IA sans autorisation demeurent une menace importante pour le secteur des médias, car ils érodent les revenus et ont une incidence sur la qualité et la diversité des médias. La Commission réexaminera

et à certains droits voisins du droit d'auteur dans le domaine de la propriété intellectuelle et la directive [2009/24/CE](#) concernant la protection juridique des programmes d'ordinateur.

⁵⁸ La première proposition de la Commission européenne remonte au 14 septembre 2016. [COM\(2016\) 593 final](#). Les études préparatoires utilisées pour la directive sur le droit d'auteur dans le marché unique numérique, qui sont encore plus anciennes, ne font pas davantage référence à l'IA. Voir Triaille J.P. et al., [Study on the legal framework of text and data mining \(TDM\)](#) (en anglais), Publication office of the European Union, mars 2014.

⁵⁹ Considérants 8 et 18 de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

⁶¹ Commission européen, [Appel à contributions](#), Initiative ciblée pour un meilleur environnement de droit d'auteur en faveur de la créativité et de l'innovation européennes..

la directive sur le droit d'auteur dans le marché unique numérique [74](#) et se penchera sur la manière d'améliorer son efficacité dans le contexte de ces évolutions. » ⁶².

Il est essentiel pour que le droit d'auteur puisse s'appliquer que les œuvres préexistantes relèvent du champ d'application des législations nationales en matière de droit d'auteur, ce qui pourrait signifier que certaines d'entre elles ne sont pas pertinentes pour l'entraînement de l'IA. Par exemple, les œuvres tombées dans le domaine public (bien que les droits moraux puissent subsister) ou les contenus qui ne sont pas considérés comme des œuvres au sens du droit d'auteur ne bénéficient pas d'une protection complète⁶³. De ce point de vue, les dispositions générales du droit d'auteur relatives à la protection des œuvres restent applicables.

2.2.1.1. Les droits concernés

Les droits ci-dessous ont été identifiés en fonction des procédures utilisées par les déployeurs d'IA au cours du processus d'entraînement, qui sont susceptibles de porter atteinte au droit d'auteur sur des œuvres préexistantes et, par conséquent, d'empêcher que l'œuvre en question ou certaines parties de celle-ci puissent être utilisées pour l'entraînement de l'IA⁶⁴ :

Table 1. Les droits d'auteur concernés par l'entraînement des systèmes d'IA

Domaine		Lien avec l'entraînement de l'IA
	Droit de reproduction	Le droit de reproduction confère aux titulaires de droits le droit exclusif d'autoriser ou d'interdire la reproduction directe ou indirecte, provisoire ou permanente, par quelque moyen et sous quelque forme que ce soit, en tout ou en partie, notamment des œuvres, des phonogrammes et des exécutions ⁶⁵ .

⁶² Communication conjointe au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, [Bouclier européen de la démocratie: renforcer la position de démocraties fortes et résilientes](#), JOIN(2025) 791 final, Bruxelles, 12 novembre 2025.

⁶³ Comme l'explique le considérant 53 de la directive sur le droit d'auteur dans le marché unique numérique, l'expiration de la période de protection d'une œuvre implique son entrée dans le domaine public et l'expiration des droits que la législation de l'Union en matière de droit d'auteur confère à l'égard de cette œuvre.

⁶⁴ [The development of Generative Artificial Intelligence from a Copyright Perspective](#) (anglais), European Union Intellectual Property Office (EUIPO), mai 2025, p. 35.

⁶⁵ Article 2 de la [directive InfoSoc](#).



Domaine		Lien avec l'entraînement de l'IA
	Droits sur les bases de données	Les bases de données bénéficient de deux types de protection : - <u>Protection au titre du droit d'auteur</u> : elle confère des droits exclusifs permettant d'autoriser la reproduction de ces bases de données, que ces reproductions soient temporaires ou permanentes⁶⁶. - <u>Droit <i>Sui generis</i></u> : il confère le droit d'interdire l'extraction et/ou la réutilisation de la totalité ou d'une partie substantielle, évaluée de manière qualitative et/ou quantitative, des contenus de cette base de données⁶⁷.
	Droits sur les programmes d'ordinateur	Dans le prolongement de la protection classique du droit d'auteur prévue par la directive InfoSoc, il existe un droit spécifique de reproduction, qu'elle soit permanente ou temporaire, ainsi que de modification, qui englobe la traduction, l'adaptation et l'arrangement⁶⁸.
	Droit des éditeurs de presse	La directive sur le droit d'auteur dans le marché unique numérique prévoit une protection spécifique sous la forme d'un droit voisin contre certaines utilisations en ligne par les fournisseurs de services de la société de l'information. Une récente interprétation de la Cour de justice de l'Union européenne autorise les législations nationales, notamment, à conférer aux éditeurs de publications de presse le droit d'obtenir une rémunération équitable en contrepartie de l'autorisation accordée aux fournisseurs de services de la société de l'information d'utiliser leurs publications, à condition : - qu'elles ne privent pas les éditeurs de publications de presse de la possibilité de refuser cette autorisation ou de l'accorder à titre gratuit ; - qu'elles n'imposent pas aux fournisseurs de services de la société de l'information une obligation de paiement sans lien avec l'utilisation de ces publications ; - que les obligations et les sanctions éventuelles imposées à ces fournisseurs respectent le principe de proportionnalité⁶⁹.

⁶⁶ Article 3 de la directive [96/9/CE](#).

⁶⁷ Article 7 de la directive [96/9/CE](#).

⁶⁸ Voir la directive [2009/24/CE](#).

⁶⁹ Voir [C-797/23](#), *op. cit.*



Domaine		Lien avec l'entraînement de l'IA
	Mesures techniques de protection	La directive InfoSoc exige que les législations nationales des États membres de l'UE prévoient une protection juridique appropriée contre le contournement de toute mesure technique efficace mise en œuvre par les titulaires de droits, pour autant que ces mesures n'empêchent pas les bénéficiaires de limitations et d'exceptions spécifiques au droit d'auteur et aux droits voisins d'en bénéficier.

Source : Observatoire européen de l'audiovisuel et EUIPO⁷⁰

2.2.1.2. Les exceptions pertinentes

Le droit d'auteur en Europe e droit d'auteur est contrebalancé par une série d'exceptions, à savoir les exceptions. Cette notion, consacrée notamment par la directive InfoSoc, vise à garantir un juste équilibre entre les droits et les intérêts des différentes catégories de titulaires de droits dans le domaine du droit d'auteur. Selon ce principe, la directive sur le droit d'auteur a introduit dans le marché unique numérique des exceptions et limitations en matière de fouille de données, de recherche scientifique, d'innovation, d'enseignement et de conservation du patrimoine culturel⁷¹.

La directive sur le droit d'auteur dans le marché unique numérique a donc instauré une exception relative à la fouille de textes et de données, en estimant que certaines organisations et institutions sont confrontées à une incertitude juridique quant à la possibilité d'effectuer des activités de fouille de textes et de données sur des contenus, et afin de garantir la légalité de cette technique lorsqu'elle est utilisée en dehors du cadre de la recherche scientifique.

En conséquence, lorsque l'utilisation de techniques de fouille de textes et de données est effectuée aussi bien par des entités privées que publiques pour analyser de grands volumes de données dans différents domaines de l'existence et à diverses finalités générales, la directive sur le droit d'auteur dans le marché unique numérique prévoit une exception en faveur de la fouille de textes et de données lorsque l'accès à l'œuvre ou à tout autre objet protégé par le bénéficiaire a été effectué de manière licite. Cette accessibilité licite vaut également lorsque l'œuvre ou tout autre objet protégé a été mis à la disposition du public en ligne, et pour autant que les titulaires de droits n'aient pas réservé de manière appropriée les droits de reproduction et d'extraction pour la fouille de textes et de données⁷².

En d'autres termes, la directive sur le droit d'auteur dans le marché unique numérique instaure un « régime de retrait » (*opt-out*), qui s'est rapidement imposé comme l'un des thèmes centraux de l'entraînement de l'IA. Quelques secteurs s'interrogent quant

⁷⁰ [The development of Generative Artificial Intelligence from a Copyright Perspective](#), *op. cit.*

⁷¹ Considérant 5 de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

⁷² Considérant 18 de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

à son application pour l'entraînement des systèmes d'IA⁷³ et à ses effets concrets sur le marché. Ces questions persistent sept ans après son adoption et quatre ans après la démocratisation de l'IA générative auprès du grand public⁷⁴.

La fouille de textes et de données est quant à elle abordée aux articles 3 et 4 de la directive sur le droit d'auteur dans le marché unique numérique. L'article 4 est en réalité au cœur du débat politique, dans la mesure où il prévoit une large exception au titre du régime de retrait (*opt-out*) en exigeant que l'exercice du retrait se fasse par l'utilisation de procédés lisibles par machine:

Table 2. Les exceptions pertinentes en matière de fouille de textes et de données dans le processus d'entraînement de l'IA

Article de la directive sur le droit d'auteur dans le marché unique numérique		Exception
	3	Exception applicable aux reproductions et extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel afin de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite.
	4	Exception ou limitation pour les reproductions et les extractions d'œuvres et d'autres objets accessibles de manière licite aux fins de la fouille de textes et de données. - L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne [régime de retrait <i>opt-out</i>].

Source : Observatoire européen de l'audiovisuel

⁷³ Dornis T. « [The Training of Generative AI is not Text and Data Mining](#) », European Intellectual Property Review (E.I.P.R.), 19 octobre 2024, en anglais. L'auteur estime que l'accent doit davantage être mis sur la violation du droit d'auteur, dans la mesure où l'entraînement de modèles d'IA générative sans licence pour les œuvres utilisées comme données d'entraînement ne relève en réalité pas du champ d'application de l'exception européenne relative à la fouille de textes et de données. Voir aussi « [Rapport Voss](#) », paragraphe K.

⁷⁴ Keller P. (IViR), [A Blind Spot at the Heart of EU Copyright and AI Policymaking?](#) (en anglais), *Kluwer Copyright Blog*, 27 mai 2026. L'auteur estime qu'outre l'incertitude autour de la mise en œuvre du régime d'opt-out, celui-ci ne parvient pas à produire les résultats escomptés car « les incitations à l'utilisation à grande échelle d'informations et de contenus accessibles au public sont façonnées par des dynamiques qui dépassent largement le champ d'application du cadre européen du droit d'auteur ».



2.2.2. La fouille de textes et de données et les obligations en matière de transparence

La directive sur le droit d'auteur dans le marché unique numérique définit la fouille de textes et de données comme toute technique d'analyse automatisée visant à analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations⁷⁵.

En matière de droit d'auteur, la fouille de textes et de données est susceptible d'impliquer des actions qui relèvent soit du droit d'auteur, soit du droit *sui generis* applicable aux bases de données, voire, dans certains cas, des deux. Plus précisément, cette situation peut se présenter lors de la reproduction d'œuvres, ou d'autres objets protégés, notamment lorsque l'extraction de contenus d'une base de données s'accompagne d'une normalisation⁷⁶ des données au cours du processus de fouille de textes et de données⁷⁷. Il en ressort une analyse automatisée des informations sous forme numérique, à savoir des textes, des sons, des images ou des données, qui relève généralement du champ d'application de la fouille de textes et de données.

Six années après la mise en place de la directive sur le droit d'auteur dans le marché unique numérique, qui encourage des déclarations de réserve de droits exprimées dans un format lisible par une machine, le « rapport Voss » estime que les systèmes actuels de réserve de droits (*opt-out*, selon le Considérant 18 CDSMD « la réservation de ces droits ne devrait être jugée appropriée que si elle est effectuée au moyen de procédés lisibles par machine »⁷⁸) sur les contenus protégés par le droit d'auteur « sont souvent peu pratiques, peuvent ne pas couvrir tous les actes pertinents d'exploration de textes et de données, et n'ont pas la transparence requise pour assurer l'effectivité de la mise en œuvre de la réglementation et du contrôle de son respect ». Le rapport reconnaît également l'existence d'éléments attestant de violations généralisées des dispositions relatives au droit d'auteur par les fournisseurs d'IA générative, parmi lesquelles le non-respect des réserves de droits en matière de fouille de textes et de données formulées par les titulaires de droits⁷⁹. L'APCE souligne par ailleurs que « [l]es exceptions relatives à la fouille de textes et de données (*text and data mining*) qui ont été adoptées avant l'avènement de l'IA générative imposent aux titulaires de droits d'auteur de s'opposer expressément à l'usage (*opt-out*) de cette exception et ne prévoient aucune disposition en matière de rémunération »⁸⁰. Néanmoins, le régime d'*opt-out* n'empêche pas l'existence des licences ou de systèmes de rémunération.

Parallèlement, la Commission européenne a rendu public en 2025 le code de bonnes pratiques sur l'IA à usage général⁸¹, qui a été signé par certains des plus importants

⁷⁵ Article 2(2) de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

⁷⁶ C'est-à-dire que les données soient organisées, nettoyées et standardisées.

⁷⁷ Considérant 8 de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

⁷⁸ Considérant 18 de la directive sur le droit d'auteur dans le marché unique numérique, *op. cit.*

⁷⁹ Considérant K du « rapport Voss ».

⁸⁰ Considérant 6 du [Résolution 2654 \(2026\)](#), La protection du droit d'auteur dans l'environnement de l'intelligence artificielle, PACE, texte adopté le 23 avril 2026.

⁸¹ Commission européenne, [General-Purpose AI Code of Practice](#) (code de bonnes pratiques pour les modèles d'IA à usage général), uniquement en anglais, publié le 10 juillet 2025.



développeurs ou fournisseurs d'IA, parmi lesquels Amazon, Anthropic, Google, IBM, Microsoft et Mistral AI. Les signataires de ce code ont reconnu⁸² l'exception ou la limitation relative à la fouille de textes et de données « qui s'applique sous réserve d'un accès licite et que l'utilisation des œuvres et autres objets protégés visés dans cet alinéa n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée [...] ».

Conformément à ce code, les signataires s'engagent à veiller à ce que seules les œuvres et autres objets protégés accessibles de manière licite puissent faire l'objet d'une reproduction et d'une extraction. Pour ce faire, ils ont pris deux engagements concernant l'utilisation de logiciels automatisés d'exploration du web (*web-crawlers*) destinés à extraire ou à compiler des données à des fins de fouille de textes et de données :

- **Ne pas contourner les mesures techniques efficaces**, telles que définies dans la directive InfoSoc, qui visent à empêcher ou à restreindre les actes non autorisés portant atteinte aux œuvres et autres objets protégés, notamment en respectant toute restriction ou tout refus d'accès technique imposé par des modèles d'abonnement ou des accès payants.
- **Exclure de leur exploration les sites web** qui mettent à la disposition du public des contenus qui, au moment de l'exploration, sont reconnus par les juridictions ou les autorités publiques de l'UE et de l'Espace économique européen comme étant constitutifs d'une violation persistante et répétée du droit d'auteur et des droits voisins à des fins commerciales.

Parallèlement, afin d'identifier et de respecter les réserves en matière de droits lors de l'exploration sur internet, conformément à l'esprit de l'article 4 de la directive sur le droit d'auteur dans le marché unique numérique, deux engagements supplémentaires de nature technique ont été pris par les signataires :

- **Utiliser des logiciels automatisés d'exploration du web** qui lisent et respectent les instructions exprimées conformément au protocole d'exclusion pour les robots (*robots.txt*), tel que spécifié dans la demande d'observations n° 9309 de l'*Internet Engineering Task Force* (IETF)⁸³.
- **Identifier et respecter d'autres protocoles appropriés pouvant être lus par des machines** pour exprimer les réserves de droits conformément à l'article 4(3) de la directive sur le droit d'auteur dans le marché unique numérique. Cette mesure peut être mise en œuvre, par exemple, au moyen de métadonnées basées sur les ressources ou sur la localisation, qui ont été adoptées par des organismes de normalisation internationaux ou européens, ou qui sont à la pointe de la technologie. Il peut s'agir de métadonnées techniquement exploitables et largement acceptées par les titulaires de droits, adaptées aux différents secteurs culturels, et faisant l'objet d'un consensus général à l'issue d'un processus inclusif mené dans le cadre de discussions de bonne foi, qui sont encouragées au niveau de l'UE.

Pour ce qui est de la transparence dans l'entraînement des IA génératives, le règlement sur l'IA impose, notamment, aux fournisseurs de systèmes d'IA – y compris les systèmes d'IA à

⁸² *Ibid.*, chapitre sur le droit d'auteur.

⁸³ Voir *Datatracker* site web, [Robots Exclusion Protocol RFC 9309](#), *op. cit.*



usage général qui génèrent des contenus audios, des images, des vidéos ou des textes synthétiques – l'obligation d'établir et de tenir à jour la documentation technique du modèle utilisé, ainsi que son processus d'entraînement et de test avec les résultats de son évaluation⁸⁴. Cette documentation doit préciser la nature et le format des données d'entrée et de sortie, par exemple s'il s'agit d'un texte ou d'une image, les spécifications techniques du modèle et du processus d'entraînement, y compris les méthodologies et techniques d'entraînement, ainsi que les choix de conception les plus pertinents, accompagnés de leur argumentation et des différentes hypothèses retenues⁸⁵. Par ailleurs, l'article 53, paragraphe 1, point d), de la CDSMD oblige les fournisseurs de modèles d'IA à usage général à établir et à mettre à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour l'entraînement du modèle d'IA à usage général.

Les fournisseurs doivent par ailleurs fournir des informations sur les données utilisées pour l'entraînement, les essais et la validation, le cas échéant, y compris le type et la provenance des données et les méthodes d'organisation, comme le nettoyage ou le filtrage, par exemple, le nombre de points de données, leur portée et leurs principales caractéristiques, la manière dont les données ont été obtenues et sélectionnées, ainsi que toutes les autres mesures visant à détecter l'inadéquation des sources de données et les méthodes permettant de déceler les biais identifiables, le cas échéant⁸⁶.

En vertu des dispositions des articles 53(2) et 54(6) du règlement sur l'IA et conformément aux lignes directrices de la Commission européenne sur la portée des obligations incombant aux fournisseurs de modèles d'IA à usage général établies par le règlement (UE) 2024/1689, des exceptions particulières s'appliquent aux fournisseurs de modèles d'IA mis à disposition dans le cadre d'une licence libre et ouverte permettant de consulter, d'utiliser, de modifier et de distribuer le modèle⁸⁷.

2.2.3. Sélection d'exemples de stratégies en matière de protection du droit d'auteur dans le domaine de l'IA

À l'instar d'autres évolutions technologiques historiques, les régimes relatifs aux droits d'auteur sont antérieurs à l'ère de l'IA, ce qui amène aujourd'hui les législateurs à s'interroger sur l'adéquation des cadres européens et nationaux en matière de droits d'auteur à l'ère de l'IA générative.

Par exemple, la récente évaluation d'impact sur le droit d'auteur et l'IA réalisée au Royaume-Uni⁸⁸ synthétise les différentes stratégies en matière de législation du droit

⁸⁴ Article 53.1.a. du règlement sur l'IA.

⁸⁵ Annexe XI du règlement sur l'IA.

⁸⁶ *Ibid.*

⁸⁷ Commission européenne, [Lignes directrices de la Commission sur la portée des obligations incombant aux fournisseurs de modèles d'IA à usage général établies par le règlement sur l'IA \(AI Act\)](#), C(2025) 7719 final, Bruxelles, 18 juillet 2025.

⁸⁸ Gov.UK, [Copyright and AI Impact Assessment](#) [Évaluation de l'impact du droit d'auteur et de l'IA] (en anglais), Ministère des Sciences, de l'Innovation et de la Technologie et ministère de la Culture, des Médias et des Sports.



d'auteur qui peuvent être observées dans le contexte technologique d'aujourd'hui, en tenant compte d'évolutions majeures telles que le cadre européen du droit d'auteur prévu par la directive sur le droit d'auteur dans le marché unique numérique.

Il n'existe à l'heure actuelle au Royaume-Uni aucune exception générale susceptible de justifier l'utilisation d'œuvres protégées par le droit d'auteur à des finalités commerciales de fouille de texte et de données ou pour l'entraînement de l'IA générative⁸⁹. À l'occasion de l'examen de la future réglementation, l'étude britannique a défini trois objectifs : accroître la transparence sur l'utilisation d'œuvres protégées par le droit d'auteur pour l'entraînement de modèles d'IA et la création de contenus générés par l'IA ; donner aux titulaires de droits sur les œuvres de création un plus grand contrôle sur l'utilisation de leur matériel pour l'entraînement de modèles d'IA et renforcer leur capacité à être rémunérés lorsque celui-ci est utilisé ; et enfin, améliorer la capacité des développeurs d'IA à accéder au matériel dont ils ont besoin pour entraîner des modèles d'IA de pointe au Royaume-Uni. En fonction de ces objectifs, quatre options se présentent généralement :

Présentée au Parlement en vertu de l'article 135 de la loi britannique relative aux données (utilisation et accès) de 2025, 18 mars 2026.

⁸⁹ House of Lords, [AI, copyright and the creative industries](#) (IA, droit d'auteur et industries créatives) – Comité des communications et du numérique, 4e rapport de la session 2024-2026 (uniquement en anglais), *Parliamentary Copyright House of Lords* 2026, 6 mars 2026.

Table 3. Entraînement de l'IA. Les différentes stratégies possibles.

0	Ne rien faire. Avec cette option, une autorisation serait généralement nécessaire pour reproduire des œuvres protégées dans les différentes étapes de l'entraînement et du développement de l'IA.
1	Renforcer le droit d'auteur, afin que des licences soient indispensables pour l'utilisation de matériel protégé par le droit d'auteur dans le cadre de l'entraînement de l'IA, et prévoir des exigences de transparence et des mesures d'accès au marché. Au Royaume-Uni, la législation relative au droit d'auteur prévoit déjà certains droits ; le système d'octroi de licences proposé dans le cadre de cette option resterait donc globalement inchangé par rapport à la situation actuelle.
2	Mettre en place une exception très étendue en matière de droit d'auteur pour la fouille de textes et de données. Cette exception ne serait pas totalement sans contrainte, puisque les activités de fouille de données ne seraient autorisées que si l'utilisateur dispose d'un accès licite. Les titulaires de droits ne pourraient toutefois pas réserver leurs droits ni se soustraire à l'exception.
3	Mettre en place une exception au droit d'auteur pour les activités de fouille de textes et de données assortie d'un mécanisme de réserve de droits et d'exigences de transparence. Cette exception s'appliquerait uniquement si un titulaire de droits n'a pas expressément réservé ses droits. Ce mécanisme permettrait ainsi aux titulaires de droits de « désactiver » l'exception pour certaines œuvres.

Source : *AI, copyright and the creative industries. Communications and Digital Committee*⁹⁰ (IA, droit d'auteur et industries créatives – Comité des communications et du numérique), 4^e rapport de la session 2024-2026, uniquement en anglais.

Comme les régimes de droit d'auteur varient à travers le monde et que les stratégies en matière d'entraînement de l'IA diffèrent en fonction des pays, on observe des situations très différentes à l'échelle mondiale lorsqu'il s'agit d'intégrer l'entraînement de l'IA dans la réglementation sur le droit d'auteur. Par exemple, des pays comme la France discutent actuellement de l'introduction d'une présomption d'utilisation des œuvres⁹¹.

Contrairement aux instruments généraux visant à harmoniser la législation en matière d'IA qui font leur apparition, comme la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle, les droits de l'homme, la démocratie et l'État de droit, il n'existe pas de convention internationale majeure consacrée à l'IA et au droit d'auteur. Ce dernier

⁹⁰ *Ibid.*, op. cit., p. 15.

⁹¹ Sénat Français, [Proposition de loi relative à l'instauration d'une présomption d'exploitation des contenus culturels par les fournisseurs d'intelligence artificielle](#), en cours, consulté le 10 juillet.

continue généralement à être réglementé par la Convention de Berne⁹², le WCT⁹³, le WPPT⁹⁴ et des instruments nationaux et européens.

Depuis 2026, plusieurs stratégies en matière d'IA et de droit d'auteur ont été proposées :

Table 4. Les stratégies européennes et nationales d'entraînement de l'IA

	EU	Les exceptions relatives à la fouille de textes et de données prévues aux articles 3 et 4 de la directive sur le droit d'auteur dans le marché unique numérique. Ce régime est le seul à disposer d'un mécanisme de droit de réserve (« <i>opt-out</i> »).
	UK	Une exception réglementaire applicable uniquement à la fouille de textes et de données à des fins de recherches non commerciales, à condition que l'utilisateur dispose d'un accès licite à l'œuvre. Des discussions politiques à ce sujet sont en cours.
	USA	Il n'existe aucun dispositif spécifiquement applicable à la fouille de textes et de données. La doctrine américaine d'une « utilisation équitable » (« <i>fair use</i> ») s'applique ⁹⁵ .
	CANADA	Il n'existe aucun dispositif spécifiquement applicable à la fouille de textes et de données. Des discussions politiques à ce sujet sont en cours ⁹⁶ .
	JAPON	De nombreuses exceptions réglementaires autorisent la fouille de textes et de données à des fins commerciales et non commerciales, et plus particulièrement à des « fins autres que le divertissement » ⁹⁷ .

⁹² Organisation Mondiale de la propriété intellectuelle (OMPI), [Convention de Berne pour la protection des œuvres littéraires et artistiques](#), adopté le 9 septembre 1886.

⁹³ [WIPO Copyright Treaty](#), adopté le 20 décembre 1996.

⁹⁴ [WIPO Performances and Phonograms Treaty](#), adopté le 20 décembre 1996.

⁹⁵ L'approche américaine en matière d'entraînement de l'IA revêt une importance particulière, compte tenu du fait que certains des principaux fournisseurs d'IA sont originaires des États-Unis. Conformément à la doctrine du *fair use*, consacrée par l'article 107 de la loi de 1976 relative au droit d'auteur, l'utilisation équitable d'une œuvre protégée par le droit d'auteur ne saurait être considérée comme une contrefaçon dès lors que certaines conditions sont remplies. Quatre critères doivent ainsi être pris en compte : 1. La finalité et la nature de l'utilisation de l'œuvre (y compris l'évaluation de la dimension commerciale) ; 2. La nature de l'œuvre protégée par le droit d'auteur ; 3. La quantité et l'importance de la partie utilisée au regard de l'œuvre protégée dans son ensemble ; 4. L'effet de cette utilisation sur le marché potentiel de l'œuvre protégée. Pour consulter la position officielle de l'Office américain du droit d'auteur, voir U.S copyright office, [Copyright and Artificial Intelligence, Partie 3, Entraînement de l'IA générative](#) (en anglais), mai 2025.

⁹⁶ Voir le Gouvernement du Canada site web, [Consultation sur le droit d'auteur à l'ère de l'intelligence artificielle générative](#), (dernière modification le 11 février 2025).

⁹⁷ [AI, copyright and the creative industries](#), *op. cit.*, p. 12.



	CHINE	Il n'existe aucun cadre spécifique. Le régime du droit d'auteur combine le système du droit d'auteur, avec notamment une utilisation équitable (<i>fair use</i>), et le système de la paternité de l'œuvre ⁹⁸ .
---	--------------	--

Source : Observatoire européen de l'audiovisuel.

Pour ce qui est de la réglementation européenne, il convient de tenir compte du fait que certaines des dispositions relatives au champ d'application territorial⁹⁹ prévues par le règlement sur l'IA précisent que ce dernier s'applique aux exploitants de systèmes d'IA dont le lieu d'établissement ou la localisation se situe dans l'Union, ainsi qu'aux fournisseurs et exploitants de systèmes d'IA dont le lieu d'établissement ou la localisation se situe dans un pays tiers, lorsque les résultats générés par le système d'IA sont utilisés au sein de l'Union¹⁰⁰.

La directive sur le droit d'auteur dans le marché unique numérique ne prévoit aucune disposition spécifique à cet égard¹⁰¹.

⁹⁸ Zhang J. et Li X., [Fair Use in Training AI Models: A Review and Prospect of the Relevant Legal Development in China](#) (en anglais), Cambridge University Press, 24 avril 2026.

⁹⁹ Considérant 106 CDSMD : « Les fournisseurs qui mettent des modèles d'IA à usage général sur le marché de l'Union devraient veiller au respect des obligations pertinentes prévues par le présent règlement. À cette fin, les fournisseurs de modèles d'IA à usage général devraient mettre en place une politique visant à respecter la législation de l'Union sur le droit d'auteur et les droits voisins, en particulier pour identifier et respecter la réservation de droits exprimées par les titulaires de droits conformément à l'article 4, paragraphe 3, de la directive (UE) 2019/790. Tout fournisseur qui met un modèle d'IA à usage général sur le marché de l'Union devrait se conformer à cette obligation, quelle que soit la juridiction dans laquelle se déroulent les actes pertinents au titre du droit d'auteur qui sous-tendent l'entraînement de ces modèles d'IA à usage général. Cela est nécessaire pour garantir des conditions de concurrence équitables entre les fournisseurs de modèles d'IA à usage général, lorsqu'aucun fournisseur ne devrait pouvoir obtenir un avantage concurrentiel sur le marché de l'Union en appliquant des normes en matière de droit d'auteur moins élevées que celles prévues dans l'Union. »

¹⁰⁰ Article 2 du règlement sur l'IA.

¹⁰¹ Zapala I., [Territorial scope of the authorization requirement and liability regime under Directive \(EU\) 2019/790 on copyright and related rights in the digital single market](#) (uniquement en anglais), *Journal of Intellectual Property Law & Practice*, volume 16, n° 8, p. 827 à 849, août 2021.



3. La question du droit d'auteur dans l'entraînement de l'IA et dans les instructions données à l'IA

3.1. Les activités d'entraînement de l'IA

Bien que l'exception relative à la fouille de textes et de données prévue par la réglementation européenne soit souvent considérée comme s'appliquant à l'entraînement des systèmes d'IA, certains s'inquiètent toutefois du champ d'application effectif de l'ensemble du processus d'entraînement au regard des articles 3 et 4 de la directive sur le droit d'auteur dans le marché unique numérique.

Même si le dispositif semble globalement satisfaisant pour le secteur audiovisuel¹⁰², la doctrine universitaire suggère que l'entraînement de l'IA va au-delà de la fouille de textes et de données, qui en constitue la partie finale, et non l'inverse, et rappelle qu'aucun instrument juridique n'a encore considéré l'entraînement de l'IA et la fouille de textes et de données comme deux activités équivalentes¹⁰³. Cette doctrine souligne en outre que la fouille de textes et de données concerne la collecte, le traitement et l'analyse d'informations, mais ne permet pas de couvrir des aspects importants de l'entraînement de l'IA, notamment ceux liés au contenu¹⁰⁴. D'un point de vue institutionnel, l'Office de l'Union européenne pour la propriété intellectuelle (EUIPO) adopte une position plus nuancée visant à concilier les différentes positions¹⁰⁵. Face à certaines incertitudes quant aux véritables effets de l'exception relative à la fouille de textes et de données, le « rapport Voss » a également récemment rappelé que le régime de retrait « *opt-out* » « peut ne pas couvrir tous les actes pertinents de la fouille de textes et de données ». Certaines décisions importantes sont rendues en Europe au niveau national. Parmi les affaires les plus pertinentes au niveau national figure celle opposant la société de gestion collective allemande GEMA à OpenAI, dans laquelle le tribunal régional de Munich a conclu que

¹⁰² Voir, par exemple, [AI Principles for Audiovisual Media Services](#) (en anglais), Bruxelles, 16 juin 2026 : « Le cadre de l'UE en matière de droit d'auteur et d'IA — en particulier la directive sur le droit d'auteur dans le marché unique numérique et le règlement sur l'IA — définit les conditions d'utilisation des œuvres protégées pour l'entraînement des systèmes d'IA. Bien que de nombreuses incertitudes subsistent quant à sa mise en œuvre, ce cadre est suffisamment flexible pour s'adapter aux évolutions technologiques et offre une base fiable pour évaluer les infractions, tant au stade des données d'entrée que des résultats générés par les systèmes d'IA ».

¹⁰³ Voir Rosati E., [The future of the movie industry in the wake of generative AI: A perspective under EU and UK copyright law](#), *op. cit.* Eleonora Rosati souligne qu'en matière de fouille de textes et de données, « les principaux actes visés par le droit britannique et européen sont ceux d'extraction, si les données sont intégrées dans une base de données, et de reproduction », et aucun autre.

¹⁰⁴ Voir Dornis T., [The Training of Generative AI is not Text and Data Mining](#), *European Intellectual Property Review* (E.I.P.R.), 19 octobre 2024.

¹⁰⁵ Voir [The development of Generative Artificial Intelligence from a Copyright Perspective](#) (en anglais), *op. cit.*, p. 67.



ChatGPT avait reproduit des paroles de chansons protégées¹⁰⁶. Dans d'autres affaires, telles que *Getty Images c. Stability AI* au Royaume-Uni, la Haute Cour britannique a proposé certains éléments d'orientation, mais les questions pertinentes n'ont toujours pas été tranchées¹⁰⁷.

L'entraînement des IA génératives peut prendre plusieurs formes et comporter différentes sous-étapes, puisqu'il s'agit d'un processus technique relativement complexe. Les derniers modèles d'IA générative sont multimodaux, ce qui leur permet de traiter divers types de contenus¹⁰⁸. En principe, ces systèmes exploitent toutes sortes d'ensembles de données, notamment des vidéos, des textes ou des images issus d'institutions privées et publiques, de particuliers ou encore de données synthétiques. L'une des méthodes les plus répandues pour accéder à ces données consiste à recourir au *web scraping* et au *crawling*¹⁰⁹. Les données ainsi acquises sont ensuite retraitées afin de garantir leur qualité, et notamment leur exactitude, leur cohérence, leur pertinence et d'autres aspects importants. Enfin, l'entraînement inclut un ajustement précis avant la génération d'un résultat. Par conséquent, du point de vue du droit d'auteur, si l'on examine les différentes étapes de l'entraînement de l'IA, les implications en matière de droit d'auteur peuvent être variables, puisque les activités qui relèvent de chacune d'entre elles sont également différentes. Selon les universitaires, l'entraînement de l'IA comporte cinq étapes pertinentes qui impliquent des processus techniques distincts :

¹⁰⁶ Par exemple, en ce qui concerne l'entraînement des IA, affaire [GEMA c. OpenAI](#). 42 O 14139/24. *Landgericht München I*.

¹⁰⁷ Stokes G. et al., [Getty Images c. Stability AI : la décision de la Haute Cour britannique apporte des éclaircissements, mais des questions essentielles sur l'IA et la violation du droit d'auteur subsistent](#) (en anglais), *DLA Piper*, 5 novembre 2025.

¹⁰⁸ Marcelin T. and Casseti F., [AI and copyright: The training of general-purpose AI](#) [IA et droit d'auteur : l'entraînement des IA à usage général], European Parliamentary Research Service, *Think Tank*, avril 2025.

¹⁰⁹ Comme l'explique l'EUIPO : « L'exploration (*crawling*) consiste généralement à parcourir de manière systématique des pages web accessibles au public afin d'indexer des métadonnées et des contenus pour les moteurs de recherche. Ce processus est généralement considéré comme moins intrusif et sert principalement à faciliter la découvrabilité. Le *scraping*, en revanche, consiste à extraire des données ou des contenus spécifiques, en ciblant souvent des informations plus granulaires et en reproduisant des parties d'œuvres d'une manière plus susceptible de soulever des préoccupations en matière de droit d'auteur ». [The development of Generative Artificial Intelligence from a Copyright Perspective](#), *op. cit.*, p. 139.

Figure 3. Le processus technique d'entraînement des systèmes d'IA.

- 
1. **Ingestion et prétraitement**, étape au cours de laquelle le contenu est collecté ;
 2. **Boucle d'entraînement**, au cours de laquelle les systèmes d'IA génèrent des données davantage affinées ;
 3. **Artefact du modèle**, où les représentations des données obtenues sont générées mais ne peuvent être visualisées ;
 4. **Génération et extraction**, étape au cours de laquelle les indications de l'utilisateur sont symbolisées ;
 5. **Journaux et caches**, dans lesquels les informations sont sélectionnées et partiellement mises en cache en vue d'un entraînement ultérieur.

Source : Copyright in Formaldehyde : How GEMA v OpenAI Freezes Doctrine and Chills AI – Partie 1 ¹¹⁰

3.2. Les instructions (*prompts*) destinées à l'IA

Bien que le terme « *prompt* » ne soit pas défini dans les instruments juridiques, il peut être entendu comme le processus qui consiste à donner des instructions ou des informations spécifiques à un système d'IA afin d'obtenir un résultat particulier, étant entendu que la clarté et la précision des instructions ou des *prompts* influenceront favorablement la qualité du résultat¹¹¹.

Les *prompts* constituent le déclencheur permettant d'obtenir un résultat dans les *chatbots* conversationnels. Outre le débat autour de la protection par le droit d'auteur des résultats, qui seront examinés dans la deuxième partie de ce rapport, pour lesquels les *prompts* apparaissent comme un outil indispensable dans l'IA générative et les IA de type *chatbot*, une deuxième interrogation concerne la protection des *prompts* en tant que tels, qui relève à la fois des dispositions générales du droit d'auteur, lesquelles ne protègent pas les simples idées, et éventuellement d'autres domaines du droit, comme celui du secret des affaires¹¹².

Aux États-Unis, le Bureau du droit d'auteur (*Copyright Office*) a conclu que « les *prompts* à eux seuls ne constituent pas un contrôle humain suffisant pour considérer les utilisateurs d'un système d'IA comme les auteurs du résultat. Les *prompts* fonctionnent principalement comme des instructions qui véhiculent des idées qui ne bénéficient

¹¹⁰ Voir Frosio G., [Copyright in Formaldehyde: How GEMA v OpenAI Freezes Doctrine and Chills AI – Part 1](#) [Le droit d'auteur sous formaldehyde : comment l'affaire GEMA c. OpenAI fige la doctrine et freine l'IA – Partie 1] (en anglais), *Kluwer Copyright Blog*, 10 décembre 2025. Voir également, pour un aperçu complet du processus d'entraînement, la pile technologique de l'IA générative, de la collecte des données à l'interaction avec l'utilisateur final, dans la [Note d'orientation du Conseil de l'Europe sur les implications de l'IA générative pour la liberté d'expression](#), CDMSI(2025)15, 3 décembre 2026, p. 44 et 45.

¹¹¹ Voir le KU Leuven site web, [Guiding AI with prompting](#).

¹¹² Pour une réflexion sur le secret des affaires et l'IA, voir Hrdy C., « [Trade Secrecy Meets Generative AI](#) », *Chicago Kent Law Review*, n° 100, 3 février 2025.



d'aucune protection. Bien que des *prompts* très détaillés puissent contenir les éléments d'expression souhaités par l'utilisateur, ils ne permettent pas, à l'heure actuelle, de contrôler la manière dont le système d'IA les traite pour générer le résultat ». Dans le même ordre d'idées, le Bureau américain du droit d'auteur a averti que « les différences entre les *prompts* et les résultats obtenus démontrent que l'utilisateur n'a pas le contrôle sur la conversion de ses idées en une expression figée, et que le système est en grande partie responsable de la sélection des éléments expressifs dans le résultat »¹¹³.

¹¹³ U.S copyright office, [Copyright and Artificial Intelligence. Part 2. Copyrightability](#) [Droit d'auteur et intelligence artificielle, Partie 2, Admissibilité au droit d'auteur] (en anglais), Janvier 2025.



4. Plateformes et utilisateurs : les conditions générales d'utilisation

La législation régit entièrement ou partiellement l'entraînement de l'IA, et les plateformes intègrent généralement certaines dispositions juridiques dans leurs conditions d'utilisation, que les utilisateurs acceptent lorsqu'ils ont recours à ces services. Dans ce contexte, les conditions d'utilisation établissent une relation privée entre l'utilisateur et la plateforme, laquelle régit plusieurs aspects tels que le droit d'auteur et l'entraînement. Toutefois, en vertu de la législation européenne, une plateforme peut dans toute situation engager sa responsabilité au titre des régimes juridiques nationaux en vigueur.

Par exemple, en matière de droit d'auteur européen, l'application de l'exception relative à la fouille de textes et de données peut engager la responsabilité d'un fournisseur d'IA en cas de violation du régime de retrait (*opt-out*). Dans d'autres cas, la responsabilité des plateformes d'IA peut être mise en cause pour violation d'autres réglementations comme le règlement sur l'IA et le règlement sur les services numériques (DSA), dans la mesure où ces services d'IA sont fournis par le biais de plateformes numériques qui peuvent relever de leur champ d'application¹¹⁴.

Les conditions d'utilisation, lorsqu'elles existent, doivent être pleinement prises en compte ; cependant, la responsabilité du fournisseur ne saurait être exclue au seul motif que ces conditions d'utilisation attribuent la responsabilité à l'utilisateur¹¹⁵.

Certaines des plateformes d'IA générative les plus populaires font en effet référence aux données saisies et à l'utilisation ultérieure que les plateformes peuvent faire des *prompts* et autres éléments fournis par les utilisateurs. Dès lors qu'il s'agit d'une acceptation des conditions d'utilisation par l'utilisateur, et que celle-ci constitue donc une relation privée et exécutoire, des implications en matière de droit d'auteur peuvent exister à l'égard des données saisies, et l'utilisateur accepte généralement que le contenu puisse être utilisé à des fins d'entraînement.

À titre d'exemple, voici plusieurs des situations les plus fréquemment observées dans les outils d'IA générative couramment utilisés :

¹¹⁴ [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). Conformément à l'article 2, le règlement sur les services numériques (*Digital Service Act* – DSA) s'applique aux services d'intermédiation proposés à des destinataires de services ayant leur lieu d'établissement ou se trouvant dans l'Union, quel que soit le lieu d'établissement des fournisseurs de ces services d'intermédiation.

¹¹⁵ Voir E. Rosati, [The future of the movie industry in the wake of generative AI: A perspective under EU and UK copyright law](#), *op. cit.*

Table 5. Dispositions relatives à l'entraînement de l'IA dans les conditions d'utilisation de certaines des plateformes d'IA les plus pertinentes.

Plateforme	Conditions d'utilisation (mentions pertinentes)
 Adobe ¹¹⁶	Vos invites et contributions aux fonctionnalités d'IA générative dans les produits Adobe, ainsi que les résultats générés par celles-ci, peuvent être examinés par des méthodes automatisées (telles que le machine learning) et manuelles à des fins de prévention des abus et de filtrage du contenu. Il est interdit d'utiliser les fonctionnalités d'IA générative d'Adobe pour créer, charger ou partager du contenu qui viole le droit d'auteur, de marque commerciale, de confidentialité, de publicité ou d'autres droits de tiers. Cela peut inclure la saisie d'invites textuelles conçues pour générer du contenu protégé par le droit d'auteur, de marque commerciale ou contrefait, le chargement d'une contribution ou d'une image de référence qui inclut du contenu protégé par le droit d'auteur d'un tiers, la génération de texte qui plagie le contenu d'un tiers, ou l'utilisation des informations personnelles d'un tiers en violation de ses droits en matière de confidentialité ou de protection des données.
 ChatGPT ¹¹⁷	Vous pouvez fournir des entrées aux Services (« Données d'Entrée ») et recevoir des sorties des Services basées sur les Données d'entrée (« Données de sortie »). Refus. Si vous ne souhaitez pas que nous utilisions votre Contenu pour entraîner nos modèles, vous avez la possibilité de vous y opposer en suivant les instructions de cet article. Veuillez noter que dans certains cas, cela peut limiter la capacité de nos Services à répondre au mieux à votre cas d'utilisation spécifique.
 Claude ¹¹⁸	C'est à vous de choisir d'autoriser ou non l'utilisation de vos données pour améliorer les nouveaux modèles Claude, et vous pouvez modifier votre choix à tout moment dans vos paramètres de confidentialité.
 Copilot ¹¹⁹	Nous ne sommes pas propriétaires de Votre contenu, mais nous pouvons l'utiliser pour exploiter Copilot et l'améliorer. En utilisant Copilot, vous nous autorisez à utiliser Votre contenu, ce qui signifie que nous pouvons le copier, le distribuer, le transmettre, l'afficher publiquement, l'exécuter publiquement, le modifier, le traduire et le reformater, et nous pouvons accorder ces mêmes droits à d'autres personnes qui travaillent en notre nom. Nous décidons d'utiliser ou non Votre Contenu, et nous n'avons pas à vous payer, à vous demander l'autorisation ou à vous dire quand nous le faisons. Mais cela ne signifie pas que nous pouvons l'utiliser comme nous le voulons. La Déclaration de confidentialité Microsoft explique comment nous utilisons

¹¹⁶ Voir le Adobe site web, [Directives à l'intention des utilisateurs de l'IA générative Adobe](#).

¹¹⁷ Voir le OpenAI site web, [Conditions d'utilisation](#).

¹¹⁸ Voir le Anthropic site web, [Updates to Consumer Terms and Privacy Policy](#) [Mises à jour des Conditions générales d'utilisation et de la Politique de confidentialité] (Entrée en vigueur le 28 août 2025).

¹¹⁹ Voir le Microsoft site web, [Conditions d'utilisation de Microsoft Copilot](#) (Entrée en vigueur le 12 juin 2026).



Plateforme	Conditions d'utilisation (mentions pertinentes)
	Votre contenu, et les options de confidentialité de Copilot vous permettent de contrôler certaines de ces utilisations.
 Midjourney ¹²⁰	En utilisant les services, vous accordez à Midjourney, à ses successeurs et à ses ayants droit une licence de droit d'auteur perpétuelle, mondiale, non exclusive, sous-licenciable sans frais, libre de redevance et irrévocable pour reproduire, préparer des travaux dérivés, afficher publiquement, exécuter publiquement, sous-licencier et distribuer des textes et des images que vous entrez dans les services, ainsi que tout actif produit par vous par le biais du service. Cette licence survit à la résiliation du présent accord par l'une ou l'autre des parties, pour quelque raison que ce soit.

¹²⁰ Voir le Midjourney site web, [Conditions d'utilisation](#) (Entrée en vigueur le 27 mai 2026).



5. Les principaux enseignements

Le développement de l'IA poursuit sa progression fulgurante, l'IA générative s'intègre de mieux en mieux dans les utilisations professionnelles et privées, et l'IA agentique connaît une forte croissance. Son impact est particulièrement visible dans les secteurs de la création, comme la production audiovisuelle, où l'IA est capable de générer du texte, des images, du son et de la vidéo. Cette évolution rapide ouvre de nouvelles perspectives, mais soulève également de complexes défis en matière de droit d'auteur, notamment au regard du cadre européen du droit d'auteur, auxquels les législateurs tentent de répondre.

L'entraînement des IA est au cœur de l'interaction entre le droit d'auteur et l'IA, puisque les modèles d'IA sont entraînés à partir de vastes ensembles de données qui peuvent inclure des œuvres protégées par le droit d'auteur et d'autres objets protégés. L'utilisation de contenus protégés par le droit d'auteur sans autorisation suscite des inquiétudes parmi les créateurs, les éditeurs et les professionnels du secteur de la culture, notamment le secteur audiovisuel, où les tâches cognitives revêtent une grande importance. D'un point de vue politique, l'accent est actuellement mis sur la recherche d'un équilibre entre l'accès aux données et la protection des titulaires de droits.

En Europe, la législation spécifique en matière de droit d'auteur applicable à l'IA générative se trouve dans certaines dispositions du règlement sur l'IA et de l'application consécutive de la directive sur le droit d'auteur dans le marché unique numérique. Le règlement sur l'IA impose des obligations de transparence, tandis que la directive sur le droit d'auteur dans le marché unique numérique traite des questions relatives au droit d'auteur. Les futures décisions de justice et les révisions législatives devraient permettre de lever les incertitudes qui subsistent.

En ce qui concerne l'entraînement de l'IA, les dispositions relatives à la fouille de textes et de données sont au cœur du débat, car elles constituent une base juridique potentielle pour certaines activités d'entraînement de l'IA. Néanmoins, des interrogations subsistent sur la manière dont ces exceptions permettent de répondre à la complexité des processus d'entraînement de l'IA, qui impliquent de multiples étapes techniques. L'efficacité des mécanismes de réserve de droits, comme le mécanisme européen de retrait (*opt-out*), fait également l'objet de discussions, comme l'a notamment souligné le récent « rapport Voss ».

Les *prompts* soulèvent également des questions relatives au droit d'auteur, notamment au sujet de la protection des prompts et de l'utilisation de contenus protégés par le droit d'auteur. Comme les *prompts* servent principalement à transmettre des instructions plutôt qu'à déterminer l'expression finale, les approches actuelles s'interrogent sur le contrôle créatif exercé par les utilisateurs sur leurs matériaux considérant les conditions des plateformes d'IA les plus populaires. La relation entre l'apport humain et le contenu généré par l'IA demeure une question juridique en constante évolution.

Une publication
de l'Observatoire européen de l'audiovisuel

