

## Proposition de loi relative à l'instauration d'une présomption d'exploitation des contenus culturels par les fournisseurs d'intelligence artificielle

Le 12 décembre 2025, une Proposition de loi (n°220) composée d'un article unique, a été enregistrée à la Présidence du Sénat. Elle vise à insérer, au sein de la section 1 du chapitre Ier du titre III du livre III de la première partie du Code de la propriété intellectuelle, un nouvel article L. 331-4-1<sup>1</sup>.

En substance, ce dispositif tend à instituer une présomption légale d'exploitation des contenus culturels par les fournisseurs de systèmes d'intelligence artificielle (ci-après « **IA** »)<sup>2</sup>.

La présente analyse revient d'abord sur les difficultés liées à l'utilisation des données-œuvres pour l'entraînement des modèles d'IA<sup>3</sup> (I), avant d'examiner la manière dont la proposition de loi entend pallier les insuffisances du cadre juridique existant par l'instauration d'une présomption d'exploitation (II). Elle analyse enfin les apports et la portée de ce mécanisme au regard des objectifs poursuivis par le législateur (III).

<sup>1</sup> Sénat, Proposition de loi « *Présomption d'exploitation des contenus culturels par les fournisseurs d'IA* », Texte n° 220 (2025-2026) de Mme Laure DARCOS et plusieurs de ses collègues, déposé au Sénat le 12 décembre 2025 : <https://www.senat.fr/leg/ppl25-220.html>

<sup>2</sup> Pour rappel, au sens du Règlement (UE) 2024/1689 sur l'IA (dit « **AI Act** »), un **fournisseur** désigne « *une personne physique ou morale, une autorité publique, une agence ou tout autre organisme qui développe ou fait développer un système d'IA ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit* » ([art. 3§3](#)) et un **système d'IA** est « *un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement, et qui, pour des objectifs explicites ou implicites, déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels* » ([art. 3§1](#)).

<sup>3</sup> La notion de **modèle d'IA** fait référence à celle de « **Modèle d'IA à usage général** » au sens de l'AI Act qui est défini comme « *un modèle d'IA, y compris lorsque ce modèle d'IA est entraîné à l'aide d'un grand nombre de données utilisant l'auto-supervision à grande échelle, qui présente une généralité significative et est capable d'exécuter de manière compétente un large éventail de tâches distinctes, indépendamment de la manière dont le modèle est mis sur le marché, et qui peut être intégré dans une variété de systèmes ou d'applications en aval, à l'exception des modèles d'IA utilisés pour des activités de recherche, de développement ou de prototypage avant leur mise sur le marché* » ([art. 3§3](#)).

Les difficultés soulevées par l'utilisation des données-œuvres à des fins d'entraînement des modèles d'IA

L'essor de l'IA a constitué un « *bouleversement d'une rare ampleur pour le secteur culturel* »<sup>4</sup>. Les modèles d'IA générative reposent en effet, pour leur entraînement, sur l'exploitation de volumes considérables de données, collectées au moyen de techniques de moissonnage automatisé (« **web scrapping** »)<sup>5</sup>. Ces données d'entraînement proviennent majoritairement de contenus numériques accessibles au public, agrégés notamment via des ensembles de données tels que le « *Common Crawl* » ou au moyen de bases d'entraînement dites « *fantômes* » (« **shadow libraries** »)<sup>6</sup>.

Or, parmi ces données figurent de nombreux contenus culturels protégés par des droits de propriété intellectuelle, susceptibles d'avoir été collectés et exploités sans l'autorisation préalable de leurs titulaires.

Cette pratique soulève au moins deux enjeux majeurs :

- D'une part, les titulaires de droits ne disposent, en pratique, d'aucune information sur l'utilisation de leurs œuvres dans le cadre de l'entraînement des modèles d'IA générative. Cette opacité, inhérente au fonctionnement de ces systèmes et fréquemment décrite comme un effet de « *boîte noire* », apparaît difficilement conciliable avec les exigences de transparence traditionnellement attachées à l'exploitation des œuvres protégées.
- D'autre part, les données-œuvres constituent un élément central de la chaîne de valeur des modèles d'IA générative, dont la gratuité semble tenue pour acquise par les fournisseurs de ces technologies. Or, sans l'accès à de vastes corpus de données culturelles, il a déjà été avancé que ces modèles seraient exposés à un phénomène de dégénérescence, compromettant la qualité et la diversité de leurs résultats<sup>7</sup>.

À ces difficultés s'ajoute le fait que les données-œuvres utilisées à des fins d'entraînement se trouvent en concurrence directe avec les données de sortie produites par ces mêmes modèles. Cette situation alimente la crainte d'un risque de substitution progressive des œuvres de l'esprit par des contenus synthétiques, susceptible de fragiliser durablement les équilibres économiques du secteur culturel<sup>8</sup>.

---

<sup>4</sup> Sénat, Proposition de loi « *Présomption d'exploitation des contenus culturels par les fournisseurs d'IA* », Texte n° 220 (2025-2026) de Mme Laure DARCOS et plusieurs de ses collègues, déposé au Sénat le 12 décembre 2025 : <https://www.senat.fr/leg/ppl25-220.html> ; Exposé des motifs, p. 3 : <https://www.senat.fr/leg/ppl25-220.pdf>

<sup>5</sup> Le moissonnage désigne une « *Extraction automatisée de contenus de sites de la toile, pratiquée en vue d'un traitement spécifique* » (Délégation générale à la langue française et aux langues de France, *Vocabulaire de l'informatique*, NOR : CTNR2212247K : <https://www.enseignementsup-recherche.gouv.fr/fr/bo/22/Hebdo21/CTNR2212247K.htm>).

<sup>6</sup> Le **Common Crawl** est une base de données publique constituée par la collecte automatisée de pages web accessibles en ligne, largement utilisée pour l'entraînement des systèmes d'IA.

Les **shadow libraries** désignent des bases de données non officielles rassemblant des œuvres protégées, diffusées sans l'autorisation des titulaires de droits.

<sup>7</sup> En effet, dans le rapport du Professeur Alexandra Bensamoun sur la mise en œuvre du règlement européen établissant des règles harmonisées sur l'intelligence artificielle (RIA), il a notamment été souligné que « *l'entraînement d'un modèle d'IA sur des données synthétiques conduit à la dégénérescence de celui-ci* » (CSPLA, Mission du CSPLA : « *IA et Transparence des données d'entraînement : publication du rapport d'Alexandra Bensamoun sur la mise en œuvre du règlement européen établissant des règles harmonisées sur l'intelligence artificielle (RIA)* », 11 décembre 2024 : <https://www.culture.gouv.fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/ia-et-transparence-des-données-d-entraînement-publication-du-rapport-d-alexandra-bensamoun-sur-la-mise-en-œuvre-du-règlement-européen-établissement>).

<sup>8</sup> Ces préoccupations ont conduit le Conseil supérieur de la propriété littéraire et artistique (CSPLA) à lancer une mission relative à la protection des contenus générés avec le recours à l'IA générative, confiée à Alexandra Bensamoun et Julie Groffe-Charrier, dont les conclusions sont attendues en juin 2026.

Ces difficultés ont trouvé un écho particulier dans les travaux parlementaires récents consacrés aux relations entre création et intelligence artificielle.

À cet égard, le rapport « *Création et IA : de la prédation au partage de la valeur* », adopté à l'unanimité par la commission compétente le 9 juillet 2025, constitue une étape structurante. Les rapporteurs y formulent huit principes directeurs visant à répondre aux déséquilibres identifiés, en mettant notamment l'accent sur la transparence des données utilisées par les fournisseurs d'IA et sur la nécessité d'assurer une rémunération des titulaires de droits<sup>9</sup>.

## Une présomption destinée à pallier les insuffisances du cadre juridique existant

Cette Proposition de loi (n°220) s'inscrit dans un contexte normatif marqué par la superposition de textes européens qui, bien que convergents dans leurs objectifs, apparaissent insuffisants pour appréhender de manière satisfaisante l'utilisation des œuvres protégées dans le cadre de l'entraînement des systèmes d'intelligence artificielle. Elle se situe ainsi à l'articulation de la Directive (UE) 2019/790 du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique (dite « **DAMUN** ») et du Règlement (UE) 2024/1689 du 13 juin 2024 sur l'intelligence artificielle (ci-après « **AI Act** »), dont elle entend combler certaines lacunes en matière d'IA.

S'agissant, en premier lieu, de la directive DAMUN, celle-ci prévoit deux exceptions au droit d'auteur en matière de fouille de textes et de données (« **text and data mining** »), transposées en droit français. À cet égard, l'article L. 122-5-3, III du Code de la propriété intellectuelle autorise, sous réserve d'un accès licite, les opérations de fouille de textes et de données quelle qu'en soit la finalité, y compris commerciale, sauf opposition exprimée par l'auteur au moyen de procédés appropriés, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne (mécanisme d'« **opt-out** »).

Depuis l'essor des modèles d'IA générative, cette exception est fréquemment invoquée par les fournisseurs d'IA pour justifier le moissonnage massif de contenus accessibles sur Internet. Toutefois, l'effectivité de l'opt-out apparaît largement compromise par l'opacité entourant les données-œuvres utilisées pour l'entraînement des modèles, privant les titulaires de droits de la possibilité d'exercer utilement leur faculté d'opposition.

Au-delà de ces difficultés pratiques, l'applicabilité même de l'exception de fouille de textes et de données aux systèmes d'IA est aujourd'hui vivement discutée. Les juridictions nationales ont d'ores et déjà adopté des positions divergentes. Le tribunal régional de Munich a ainsi écarté son application dans un cas de mémorisation de contenus protégés par un système d'IA, estimant que ce processus ne relevait pas de la fouille de données<sup>10</sup>. A l'inverse, le Royaume-Uni a retenu son applicabilité dans une autre affaire, illustrant l'absence d'interprétation harmonisée et l'insécurité juridique qui en résulte pour l'ensemble des acteurs concernés<sup>11</sup>.

---

CSPLA, Mission du CSPLA : « *Le CSPLA lance une mission relative à la protection des contenus générés avec le recours à l'IA générative* », 20 juin 2025 : <https://www.culture.gouv.fr/nous-connaître/organisation-du-ministère/Conseil-superieur-de-la-propriété-littéraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/le-cspla-lance-une-mission-relative-a-la-protection-des-contenus-générés-avec-le-recours-a-l-ia-générative>

<sup>9</sup> A. Evren, L. Darcos, P. Ouzoulias, au nom de la commission de la culture, de l'éducation et de la communication et du sport, « *Création et IA : de la prédation au partage de la valeur* », rapport d'information n°842 (2024-2025), déposé le 9 juillet 2025 : <https://www.senat.fr/rap/r24-842/r24-842.html>

<sup>10</sup> Tribunal de Munich, 11 nov. 2025, 42 O 14139/24, GEMA c/ Open AI : <https://www.gesetze-bayern.de/Content/Document/Y-300-Z-GRURRS-B-2025-N-30204>

<sup>11</sup> High Court of England and Wales, 4 nov. 2025, IL-2023- 000007, Getty Images v Stability AI : <https://www.judiciary.uk/wp-content/uploads/2025/01/Getty-Images-and-others-v-Stability-AI-14.01.25.pdf>

En pratique, les conditions posées pour la mise en œuvre de l’exception sont rarement réunies. D’une part, les bases d’entraînement des systèmes d’IA intègrent fréquemment des œuvres protégées, voire des contenus obtenus de manière illicite. D’autre part, les titulaires de droits se heurtent à des obstacles structurels pour exercer leur opt-out : méconnaissance de l’utilisation de leurs œuvres, absence de mécanismes standardisés et aisément accessibles d’opposition, ainsi que contournement possible des mesures techniques mises en place.

L’AI Act n’apporte qu’une réponse partielle à ces difficultés. Son article 53§1 (d) impose aux fournisseurs de modèles d’IA à usage général d’élaborer et de mettre à la disposition du public « *un résumé suffisamment détaillé du contenu utilisé pour entraîner le modèle d’IA à usage général, conformément à un modèle fourni par le Bureau de l’IA* ». A cette fin, la Commission européenne a publié, le 24 juillet 2025, un avis explicatif accompagné d’un modèle de résumé destiné à guider les fournisseurs dans l’identification des sources d’entraînement de leurs modèles<sup>12</sup>.

Si ce mécanisme contribue indéniablement à renforcer les exigences de transparence<sup>13</sup>, il ne permet toutefois ni d’identifier précisément les œuvres exploitées, ni d’assurer, à lui seul, une effectivité du droit d’opposition ou une rémunération des titulaires de droits. C’est dans cet espace normatif résiduel que s’inscrit la proposition de loi, laquelle entend instaurer une présomption d’exploitation afin de rééquilibrer les rapports entre fournisseurs d’IA et ayants droit.

## Les apports et la portée de la nouvelle présomption

L’instauration d’une présomption légale d’exploitation poursuit un double objectif, à la fois probatoire et incitatif, et s’inscrit dans une logique de rééquilibrage des rapports entre titulaires de droits et fournisseurs de systèmes d’IA.

En premier lieu, cette présomption vise à restaurer l’effectivité des droits de propriété intellectuelle en allégeant la charge de la preuve pesant sur les titulaires de droits. En pratique, la démonstration d’une exploitation non autorisée des œuvres se heurte aux difficultés inhérentes au fonctionnement des modèles d’IA générative, marqués par un effet de « *boîte noire* » et par un déficit structurel de transparence quant aux données d’entraînement utilisées.

Dans ce contexte, la présomption permettrait de tenir pour vraisemblable l’utilisation des données-œuvres d’un auteur lorsque le système d’IA générative produit un contenu « *à la manière de* » ou « *dans le style de* »

---

<sup>12</sup> Commission européenne, “*Explanatory Notice and Template for the Public Summary of Training Content for general purpose AI models*”, Policy and legislation, 24 juillet 2025 : <https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>

Il convient également de relever que le CSPLA a publié, le 11 décembre 2024, un rapport consacré à la mise en œuvre de l’AI Act, visant notamment à proposer un modèle de résumé suffisamment détaillé.

CSPLA, Mission du CSPLA : « *IA et Transparence des données d’entraînement : publication du rapport d’Alexandra Bensamoun sur la mise en œuvre du règlement européen établissant des règles harmonisées sur l’intelligence artificielle (RIA)* », 11 décembre 2024 : <https://www.culture.gouv.fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/ia-et-transparence-des-donnees-d-entraînement-publication-du-rapport-d-alexandra-bensamoun-sur-la-mise-en-œuvre-du-règlement-européen-établissement>

<sup>13</sup> Pour plus d’informations : Racine, « *Publication du modèle de résumé « suffisamment détaillé » des données d’entraînement des modèles d’IA à usage général* », juillet 2025 : <https://www.racine.eu/wp-content/uploads/2025/07/250728-flash-actu-publication-du-modele-de-resume-des-donnees-d-entraînement-des-modeles-dia-a-usage-general-v2.pdf>

celui-ci, ou encore lorsqu'il reproduit des extraits ou des éléments identifiables d'une œuvre ou d'un objet protégé, tel qu'un article de presse. Cette vraisemblance pourrait être confortée par des éléments techniques complémentaires, notamment des expertises techniques ou des travaux scientifiques portant sur le fonctionnement des modèles concernés ou sur leur chaîne d'approvisionnement en données.

En second lieu, la présomption revêt une fonction dissuasive et incitative. Elle tend à décourager certaines pratiques d'entraînement fondées sur le moissonnage massif et non maîtrisé de contenus protégés, tout en incitant les fournisseurs de systèmes d'IA à adopter des comportements plus vertueux, tels que la traçabilité des données utilisées, la mise en place de mécanismes contractuels ou la négociation de licences avec les titulaires de droits. À cet égard, le dispositif projeté s'inscrit dans une logique de responsabilisation des acteurs de l'IA et de rééquilibrage du rapport de force économique au profit des ayants droit.

\*\*\*

**En définitive, si cette proposition de loi est largement perçue par les acteurs culturels comme un instrument de rééquilibrage des rapports entre titulaires de droits et utilisateurs de technologies d'IA, elle suscite des interrogations quant à ses effets potentiels sur la compétitivité de la filière française de l'intelligence artificielle, la sécurité juridique des acteurs et la dynamique d'innovation.**

**L'issue des débats parlementaires permettra de déterminer si le législateur entend consacrer la présomption d'exploitation comme un nouvel outil structurant de régulation des usages de l'IA, ou s'il privilégiera des mécanismes alternatifs, reposant notamment sur la transparence, la contractualisation ou la négociation collective.**

## AUTEURS



**Charles Bouffier**  
Avocat associé  
[cbouffier@racine.eu](mailto:cbouffier@racine.eu)  
(+33) 6 64 10 87 36



**Naomi Meynlé-Hamza**  
Juriste-Doctorante  
[nmeynle-hamza@racine.eu](mailto:nmeynle-hamza@racine.eu)



**Océane Colas**  
Stagiaire élève-avocate  
[ocolas@racine.eu](mailto:ocolas@racine.eu)