

ASSEMBLÉE NATIONALE

4 juin 2026

RELATIVE À L'INSTAURATION D'UNE PRÉSUMPTION D'EXPLOITATION DES
CONTENUS CULTURELS PAR LES FOURNISSEURS D'INTELLIGENCE ARTIFICIELLE -
(N° 2864)

Commission	
Gouvernement	

N° 24

AMENDEMENT

présenté par
M. Bothorel

ARTICLE UNIQUE

Substituer aux alinéa 2 à 4 les deux alinéas suivants :

« L'article L. 122-3 est complété par une phrase ainsi rédigée :

« La mémorisation dans les modèles de langage d'intelligence artificielle générative peut être considérée comme une reproduction. »

EXPOSÉ SOMMAIRE

Cet amendement tire toutes les conséquences du jugement, certes non définitifs de la Cour régionale supérieure de Munich – I 42 O 14139/24 - Arrêt GEMA c. Open AI.

Dans son jugement rendu, la 42e Chambre civile du Tribunal régional de Munich I, spécialisée en droit d'auteur, a essentiellement fait droit aux demandes en cessation, en information et en dommages-intérêts formulées par la GEMA à l'encontre de deux sociétés du groupe Open AI. La société de gestion collective a soutenu que les paroles des chansons étaient mémorisées dans les modèles de langage des défendeurs et, lors de l'utilisation du chatbot, étaient restituées presque inchangées en réponse à des requêtes simples des utilisateurs.

OpenAI avait contesté, arguant que leurs modèles ne stockaient ni ne copiaient les données d'entraînement spécifiques, mais reflétaient dans leurs paramètres ce qu'ils avaient appris à partir de l'ensemble des données d'entraînement. Les sorties générées l'étant uniquement à la suite des entrées utilisateur (prompts), ce serait l'utilisateur, en tant que créateur, et non OpenAI, qui serait responsable des sorties. En tout état de cause, les éventuelles atteintes au droit

d'auteur seraient couvertes par les exceptions légales, notamment celle relative à l'extraction de textes et de données (text and data mining).

Le jugement a donc distingué le « chalutage » des données qui est légal, couvert par le Code de la propriété intellectuelle et qui n'est nullement changé dans la proposition de loi, telle qu'elle a été adoptée par la Commission des Affaires culturelles et de l'éducation. Par contre, la mémorisation dans les modèles de langage, tout comme la reproduction des paroles de chansons dans les sorties du chatbot, constituent des atteintes aux droits d'exploitation du droit d'auteur. Ces actes ne sont pas couverts par les exceptions, notamment celle relative au text and data mining.

Le tribunal a donc affirmé :
· Les exceptions TDM autorisent la préparation du corpus (phase 1).
· Elles ne couvrent pas la mémorisation dans les paramètres du modèle (phase 2).
· La mémorisation excède donc la finalité du TDM : pas d'exception applicable.
C'est sur cette base que doivent être poursuivis les propriétaires des modèles d'AI, dans le cadre de la réglementation actuelle et en conservant la logique du « opt-out » et non pas sur la base de la proposition hasardeuse de la proposition de loi telle qu'adoptée par la Commission des affaires culturelles et de l'éducation.

C'est pourquoi cet amendement propose de rester strictement dans le cadre juridique européen existant, de faire perdurer les possibilités du text and data mining, pour entraîner les modèles d'IA, et singulièrement les modèles européens qui ont quelque retard en la matière. Par contre, la mémorisation et donc la restitution au sein des modèles relèvent de l'article L122-3 du Code de la propriété intellectuelle. La précision qui est apportée par cet amendement est de notre point de vue superfétatoire, mais elle peut aider à comprendre la manière dont les juridictions pourraient appliquer :
· l'article L. 122-3 du Code de la propriété intellectuelle, qui définit la reproduction,
· les exceptions de text and data mining (TDM) prévues aux articles L. 122-5, 10° et 11° CPI.

Cette modification de l'article L123 offrirait ainsi des repères utiles pour les praticiens confrontés à la qualification juridique des opérations d'entraînement, de mémorisation et de fonctionnement des modèles d'IA générative. Elle contribuerait également à stabiliser les acquis de la jurisprudence française en matière de droit d'auteur et intelligence artificielle, sans se lancer dans la réécriture des principes du droit avec des novations telle que la présomption de culpabilité, l'inversion de la charge de la preuve, la rétroactivité. Ces forts principes étant proposés par le Sénat, le Conseil d'État et la Commission des affaires culturelles et de l'éducation Elle fournirait des éléments utiles aux praticiens confrontés à la qualification juridique des opérations d'entraînement et de fonctionnement des modèles d'IA.

Cet amendement reste un appel à poursuivre et accélérer le travail au niveau européen et à ne pas bouleverser le seul cadre français. Son adoption pousserait à un travail plus sérieux de réécriture dans le cadre de la navette parlementaire et en ayant pris soin de notifier à la Commission européenne cette volonté de modification, quand bien même le Conseil d'État estime que cela n'est

pas

nécessaire.

Analyse plus détaillée du jugement

1. Mémorisation des paroles dans les modèles de langage

Des recherches en informatique ont démontré que des données d'entraînement peuvent être intégrées dans les modèles de langage et extraites sous forme de sorties. Ce phénomène est appelé mémorisation. Il se produit lorsque les modèles de langage ne se contentent pas d'extraire des informations de l'ensemble des données d'entraînement pendant la phase d'apprentissage, mais intègrent également ces données dans les paramètres du modèle après l'entraînement. La mémorisation a été établie en comparant les paroles des chansons présentes dans les données d'entraînement avec les reproductions dans les sorties. Compte tenu de la complexité et de la longueur des paroles, une coïncidence comme cause de leur reproduction peut être exclue. La mémorisation signifie que les paroles sont fixées sous forme de données dans les paramètres du modèle, ce qui remplit la condition de reproduction au sens du droit d'auteur. Selon l'article 2 de la directive InfoSoc, une reproduction est considérée comme existante « de quelque manière et sous quelque forme que ce soit ». Le fait que les paroles soient encodées sous forme de valeurs de probabilité n'a pas d'incidence à cet égard. Les nouvelles technologies, comme les modèles de langage, sont donc couvertes par le droit de reproduction (art. 2 de la directive InfoSoc et § 16 de la loi allemande sur le droit d'auteur, UrhG). Selon la jurisprudence de la Cour de justice de l'Union européenne, une perception indirecte suffit pour caractériser une reproduction, dès lors que l'œuvre peut être perçue à l'aide d'outils techniques.

2. Absence de couverture par les exceptions légales

La reproduction dans les modèles n'est pas couverte par les dispositions relatives au text and data mining : art. L. 122-5, 10° et 11° CPI, issue de l'ordonnance n° 2021-580 du 12 mai 2021. Bien que les modèles de langage soient en principe couverts par le champ d'application des exceptions pour le text and data mining, celles-ci ne s'appliquent qu'aux reproductions nécessaires lors de la constitution du corpus de données pour l'entraînement, comme la conversion d'une œuvre dans un autre format (numérique) ou son stockage en mémoire de travail. La logique sous-jacente est que ces reproductions ne sont réalisées qu'à des fins d'analyse ultérieure et n'affectent donc pas les intérêts d'exploitation des auteurs. Comme ces actes purement préparatoires pour le text and data mining n'affectent pas les intérêts d'exploitation, la loi ne prévoit aucune obligation de rémunération pour l'auteur.

En revanche, lorsque, comme en l'espèce, non seulement des informations sont extraites des données d'entraînement, mais que des œuvres sont reproduites, cela ne relève plus du text and data mining. Le postulat du text and data mining et des exceptions associées — à savoir que l'évaluation automatisée de simples informations n'affecte pas les intérêts d'exploitation — ne s'applique pas dans cette configuration. Au contraire, les reproductions dans le modèle portent atteinte aux droits d'exploitation des ayants droit. Même si l'on devait supposer une lacune réglementaire non intentionnelle (le législateur n'ayant pas anticipé la mémorisation et la reproduction permanente dans les modèles, pertinente en droit d'auteur), il n'existe pas d'intérêt comparable. La disposition d'exception régit la licéité des actes

préparatoires de reproduction dans le cadre du text and data mining, situation dans laquelle les intérêts d'exploitation des auteurs ne sont pas menacés, car seules des informations sont extraites et l'œuvre en tant que telle n'est pas reproduite. En revanche, dans le cas des reproductions dans le modèle, l'exploitation de l'œuvre est durablement entravée et les intérêts légitimes des ayants droit sont lésés. Une application par analogie de la disposition d'exception, qui ne prévoit pas de rémunération pour l'exploitation, laisserait donc les auteurs et ayants droit sans protection. Selon ce jugement, les défendeurs ont illégalement reproduit et mis à disposition du public les paroles des chansons litigieuses via les sorties du chatbot. Les éléments originaux des paroles restaient toujours reconnaissables dans les sorties. Et ce sont les propriétaires des modèles d'AI (et non les systèmes), et non les utilisateurs, qui en sont responsables. Les sorties étaient générées par des prompts simples. Les propriétaires des modèles AI (et non les systèmes) exploitent les modèles de langage pour lesquels les paroles des chansons ont été sélectionnées comme données d'entraînement et avec lesquels ils ont été entraînés. Ils sont responsables de l'architecture des modèles et de la mémorisation des données d'entraînement. Ainsi, les modèles de langage exploités par les propriétaires des modèles d'AI ont eu une influence significative sur les sorties, et le contenu spécifique des sorties a été généré par les modèles de langage. L'atteinte aux droits d'exploitation par les sorties n'est pas non plus couverte par une exception légale.

Textes	applicables	:
· L'article	L. 122-4	CPI
· La directive (UE) 2019/790 (« DAMUN ») prévoit deux exceptions de text and data mining (TDM) par les articles 3 et 4 (TDM scientifique et TDM commercial), transposées en France (art. L. 122-5, 10° et 11° CPI, issue de l'ordonnance n° 2021-580 du 12 mai 2021).		
· Articles 2, 3 de la directive		InfoSoc
· Article 4 de la directive DSM		