

TRIBUNAL RÉGIONAL DE HAMBOURG 27 SEPTEMBRE 2024 N°310 O 227/23

1^{ère} décision sur l'exception de fouille de textes et de données « LAION c. Robert Kneschke »

MOTS CLÉS : Propriété intellectuelle – Droit d'auteur – Protection juridique des contenus numériques – Directive 2001/29/CE – Directive 2019/790 – Fouille de textes et de données – Recherche scientifique – Intelligence artificielle générative.

Le tribunal régional de Hambourg a établi un précédent en matière d'exception au droit d'auteur, interprétant l'exception de fouille de textes et de données dans le cadre de l'intelligence artificielle générative. Cette décision relance le débat portant sur l'application de la propriété intellectuelle aux innovations technologiques.

FAITS : En l'espèce « LAION » (Large-scale Artificial Intelligence Open Network) une organisation allemande à but non lucratif, a mis à disposition gratuitement un set de données d'entraînement « LAION-5B » qui a notamment été utilisé pour entraîner des logiciels d'intelligence artificielle générative. Ce set comportait un lien hypertexte vers une image sur « Bigstockphoto », un site internet que le photographe R. Kneschke utilisait pour promouvoir et vendre ses images.

PROCÉDURE : Le photographe agit en justice pour violation du droit de reproduction au motif que LAION a téléchargé une copie d'une photographie alors que les conditions d'utilisation du site internet interdisaient l'utilisation des images par des « programme automatisés ». LAION a invoqué l'exception de fouille de texte et de données à des fins de recherches scientifiques pour justifier l'utilisation de l'image protégée dans l'ensemble de données à des fins d'entraînement et développement des algorithmes d'une intelligence artificielle.

PROBLÉMATIQUE : La question qui se pose est celle de savoir si l'exception de fouille de texte et de données peut être appliquée aux modèles d'intelligence artificielle génératifs et ainsi écarter la protection du droit d'auteur sur les contenus numériques à des fins de recherches scientifiques ?

SOLUTION : Le tribunal régional de Hambourg a tout d'abord estimé que les reproductions réalisées par l'organisation LAION étaient effectuées à des fins d'exploration de textes et de données au sens de l'article 44b UrhG et à des fins de recherche scientifique dans un but scientifique au sens de l'article 60d UrhG. Il a jugé qu'il n'y avait pas de violation du droit d'auteur concernant la reproduction non autorisée de la photographie, l'utilisation étant couverte par l'exception relative à la fouille de textes et de données. Néanmoins la juridiction condamne l'organisation au motif de l'irrespect des conditions générales du site internet qui interdisait l'utilisation de la photographie par des programmes automatisés.



Sources :

- **Article 3¹ de la directive DAMUN** autorise « *les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue 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* ».
- **Article 4² de la directive DAMUN** autorise « *les reproductions et les extractions et d'autres objets protégés accessibles de manière licite aux fins de fouille de textes et de données pour tout bénéficiaire, sans restriction concernant l'objet des activités réalisées, mais en introduisant une limite par la possibilité donnée aux titulaires de droits de s'y opposer* ».
- **Article L. 122-5-3 du Code de la propriété intellectuelle** « *la mise en œuvre d'une technique d'analyse automatisée de textes et données sous forme numérique afin d'en dégager des informations, notamment des constantes, des tendances et des corrélations* ».
- **Dalloz actualité.** Première décision en Europe sur l'exception de fouille de textes et de données : l'affaire *LAION c. Robert Kneschke* – Elodie Migliore, Doctorante au CEIPI, Université de Strasbourg – 15 octobre 2024.

¹ Section 60d UrhG de la loi allemande sur le droit d'auteur

² Section 44b UrhG de la loi allemande sur le droit d'auteur



Une décision pionnière appliquant l'exception de fouille aux intelligences artificielles génératives

Le tribunal a mobilisé pour la première fois les articles 3 et 4 de la Directive DAMUN³ relatifs aux exceptions de fouille de textes et de données dans le cadre de l'entraînement des intelligences artificielles.

La juridiction a distingué 3 phases dans le processus d'entraînement : la création d'un jeu de données, son utilisation pour l'entraînement du modèle et la génération de nouveaux contenus. Le litige portait sur la première phase du processus. Selon le tribunal, la reproduction de la photographie pour créer un *dataset* était légale en vertu de l'exception de fouille, compte tenu de la mise à disposition gratuite au public, il s'agissait d'une activité scientifique, sans but commercial.

La reproduction de la photographie a été jugée compatible avec les objectifs du Règlement sur l'intelligence artificielle⁴ qui précise que la création d'un *dataset* destiné à l'entraînement des systèmes automatisés relève de l'article 4 de la Directive DAMUN. Ainsi l'autorisation préalable de l'auteur pour justifier la phase d'entraînement n'était pas nécessaire.

La juridiction s'est aussi prononcée sur la compatibilité de l'exception en l'espèce avec le test des 3 étapes⁵. L'exception doit être limitée à un usage spécifique, ne doit ni nuire à l'exploitation normale de l'œuvre ni créer un préjudice injustifié aux droits de l'auteur.

En raison de l'usage de la photographie pour l'analyse spécifique d'images et son inclusion dans un *dataset*, et à défaut de preuve d'un impact négatif de l'utilisation de l'œuvre, le *scraping* opéré par LAION pouvait être considéré comme une opération de fouille de textes et de données à des fins de recherches scientifiques, sans violation du droit d'auteur.

³ Directive 2019/790 du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique

⁴ Article 54 du règlement sur l'Intelligence artificielle 2024/1689 du 13 juin 2024

⁵ Test issu de la Directive du 22 mai 2001 2001/29/CE sur « l'harmonisation de certains aspects du droit

Enfin, la question de l'application de l'exception de reproduction temporaire a été soulevée, mais rejetée, la copie n'étant ni transitoire ni fortuite.

Le développement de dispositifs techniques comme rempart aux excès des intelligences artificielles génératives

Bien que LAION ait licitement téléchargé une copie de la photographie pour la reproduire dans un set de données, elle a été condamnée pour avoir violé les conditions d'utilisation de l'image. Conformément à l'article 4 de la Directive DAMUN, un titulaire de droit peut s'opposer à la fouille de son œuvre et à toute extraction par des « *procédés lisibles par machine* ». En l'espèce le site « *Bigstockphoto* » interdisait l'utilisation de la photographie par des programmes automatisés via un système « *opt-out* » inséré dans les conditions d'utilisation en « *langage naturel* ». LAION a contesté la réservation sous prétexte qu'elle était reproduite sous la forme d'un texte et non d'un procédé technique. Cependant, le tribunal a jugé la réservation conforme à la Directive et formulée de manière lisible par machine.

La réservation doit être appréciée « à la lumière des développements techniques existant au moment de l'utilisation de l'œuvre », car les logiciels d'intelligence artificielle sont désormais capables de « *saisir le contenu d'un texte écrit en langage naturel* ». De plus la juridiction a insisté sur l'obligation des fournisseurs d'intelligence artificielle de mettre en œuvre des stratégies efficaces pour identifier et respecter les réserves revendiquées par les auteurs⁶.

Enfin, la Directive permet aux auteurs d'envisager des réservations de droit par accord contractuel ou déclaration unilatérale.

d'auteur et des droits voisins dans la société de l'information » repris par l'arrêt rendu par la Cour de Cassation « *Mulholland Drive* » du 28 février 2006 qui précise la nature juridique des exceptions

⁶ Règlement européen sur l'Intelligence artificielle 2024/1689 du 13 juin 2024



Toutefois, ces dispositifs de protection exigent une transparence des plateformes afin de garantir l'exercice effectif des droits.

Une piste envisageable serait alors de qualifier les plateformes fournissant des modèles d'intelligence artificielle générative, d'hébergeurs assujettis aux obligations contraignantes du DSA⁷.

Le rejet d'un droit d'auteur exclusif au profit de la liberté de création et d'innovation

À l'origine, la philosophie européenne, telle qu'exprimée dans la Directive du 22 mai 2001 sur l'harmonisation du droit d'auteur, visait à valoriser les œuvres de l'esprit en protégeant l'empreinte de la personnalité de l'auteur.

Cependant cette vision a parfois entravé l'utilisation numérique des œuvres protégées et l'essor des systèmes automatisés. La Directive DAMUN a donc adapté le droit d'auteur aux évolutions technologiques pour favoriser un marché numérique, conciliant les intérêts des auteurs et l'intérêt général. L'ère du numérique est à la dilution du droit d'auteur, la réglementation européenne permet l'utilisation d'extraits d'œuvres par des systèmes d'intelligences au risque de porter atteinte au droit d'auteur.

L'article 3 de la Directive DAMUN interprète largement la notion de « *recherche scientifique* », l'exception de fouille est ouverte à toute entité, privée ou publique, sans restriction d'usage. Et, la seule mise à disposition publique et gratuite (sans objectif commercial) permet le bénéfice de l'exception de fouille, sans considération pour l'utilisation ultérieure.

Bien que la création issue de la reproduction de la photographie ne génère encore aucun gain de connaissance immédiat, cette éventualité est privilégiée permettant à des chercheurs d'en bénéficier (enrichissement commun, accès au patrimoine culturel, développement de nouvelles technologies, critères associés à la recherche scientifique). Cette interprétation de la mise à disposition à des fins scientifiques suffit à justifier l'absence de violation du droit d'auteur.

Néanmoins, l'éventuelle utilisation commerciale ultérieure par des sociétés commerciales est un risque évident pour le droit d'auteur, la gestion de l'intelligence artificielle étant potentiellement contrefaisante en l'absence du consentement de l'auteur à l'exploitation lucrative de son œuvre.

Enfin, la légitimité des systèmes d'intelligence artificielle générative, qui exploitent des contenus protégés pour satisfaire les utilisateurs, reste un enjeu majeur.

Alexia RAMOS

Master 2 Droit des Communications Electroniques

AIX-MARSEILLE UNIVERSITE, LID2MS-IREDIC 2024

⁷ Digital Service Act, règlement européen sur les services numériques, du 19 octobre 2022



ARRÊT : TRIBUNAL RÉGIONAL DE HAMBOURG 27 SEPTEMBRE 2024 N°310 O 227/23

[...] La partie défenderesse peut invoquer la prescription de l'article 60d UrhG en ce qui concerne la reproduction en cause. En conséquence, les reproductions pour l'exploration de textes et de données sont autorisées à des fins de recherche scientifique par des organismes de recherche. La reproduction a été effectuée à des fins d'exploration de textes et de données au sens de l'article 44b (1) UrhG. Elle a également été effectuée à des fins de recherche scientifique au sens de l'article 60d (1) UrhG.

La recherche scientifique désigne généralement la poursuite méthodique et systématique de nouvelles connaissances. La notion de recherche scientifique, en ce qu'elle permet déjà de considérer comme suffisant l'"effort" méthodique et systématique d'acquisition de nouvelles connaissances, ne doit pas être comprise de manière si étroite qu'elle ne couvrirait que les étapes de travail directement liées à l'acquisition de connaissances ; il suffit au contraire que l'étape de travail en question vise à une acquisition (ultérieure) de connaissances, comme c'est le cas, par exemple, des nombreuses collectes de données qui doivent d'abord être effectuées pour pouvoir ensuite tirer des conclusions empiriques. En particulier, le concept de recherche scientifique ne présuppose pas un succès ultérieur de la recherche.

Ainsi, la création d'un ensemble de données du type en cause, qui peut servir de base à l'entraînement de systèmes d'intelligence artificielle, peut assurément être considérée comme de la recherche scientifique au sens précité. Bien que la création de l'ensemble de données en tant que tel ne soit pas encore associée à un gain de connaissances, il s'agit d'une étape de travail fondamentale dans le but d'utiliser l'ensemble de données à des fins d'acquisition de connaissances à une date ultérieure. On peut affirmer qu'un tel objectif existait également dans le cas présent. Il suffit que l'ensemble de données ait été - indiscutablement - publié gratuitement et donc mis à la disposition des chercheurs (également) dans le domaine des réseaux neuronaux artificiels. Le fait que l'ensemble de données soit également utilisé par des entreprises commerciales pour l'entraînement ou le développement de leurs systèmes d'intelligence

artificielle n'est pas pertinent, car la recherche des entreprises commerciales reste également de la recherche.

Dans ce contexte, la question litigieuse entre les parties de savoir si la défenderesse effectue également de la recherche scientifique sous la forme du développement de ses propres modèles d'IA ainsi que de la création d'ensembles de données correspondants n'est pas pertinente. Le défendeur ne poursuit pas non plus de fins commerciales au sens de l'article 60d, paragraphe 2, point 1, de l'UrhG.

[...] Selon le test en trois étapes ce test, les exceptions standardisées ne peuvent être appliquées que dans certains cas particuliers où il n'est pas porté atteinte à l'exploitation normale de l'œuvre ou autre objet protégé et où les intérêts légitimes du titulaire du droit ne sont pas indûment lésés. Ces conditions sont remplies en l'espèce. La reproduction relevant du droit d'auteur en l'espèce est limitée à l'analyse des fichiers d'images en vue de leur conformité avec une description d'image préexistante et de leur enregistrement ultérieur dans un ensemble de données. Il n'apparaît pas et le demandeur ne prétend pas que cette utilisation porterait atteinte aux possibilités d'utilisation des œuvres concernées.

[...] La Chambre tend à considérer une réserve d'utilisation rédigée uniquement en "langage naturel" comme "compréhensible par une machine". Toutefois, la question de savoir si et dans quelles conditions une réserve déclarée en "langage naturel" peut également être considérée comme "compréhensible par une machine" devra toujours être résolue en fonction du développement technique existant au moment de l'utilisation de l'œuvre.

En conséquence, le législateur européen a également stipulé dans le cadre du règlement sur l'IA que les fournisseurs de modèles d'IA doivent mettre en place une stratégie, en particulier pour identifier et respecter une réserve juridique émise "également au moyen de technologies de pointe". Or, ces "technologies de pointe" comprennent sans ambiguïté les applications d'IA capables de reconnaître le contenu d'un texte écrit en langage naturel.



Ramos Alexia



Cette création par LID2MS-IREDIC est mise à disposition selon les termes de la licence [Creative Commons Paternité - Pas d'Utilisation Commerciale - Pas de Modification 2.0 France](https://creativecommons.org/licenses/by-nc-nd/2.0/fr/).