

## INTELLIGENCE ARTIFICIELLE ET DROIT D'AUTEUR

### POSITION du gf2i

Mai 2024

- Le gf2i appelle la France à maintenir sa position protectrice du droit d'auteur dans son action en faveur du développement des technologies d'intelligence artificielle.
- Le gf2i soutient toute mesure favorisant la mise en place d'un marché de la donnée pour le déploiement des outils d'intelligence artificielle.
- Le gf2i encourage les pouvoirs publics à garantir aux ayants droit un cadre juridique leur permettant de garder la maîtrise de l'exploitation des œuvres dans ce nouveau marché.
- Le gf2i plaide pour un opt-out le plus large possible, condition *sine qua non* à la mise en place de ce marché pour les contenus protégés.
- Le gf2i soutient le développement d'outils collectifs de contrôle des usages d'œuvres pour l'entraînement des systèmes d'intelligence artificielle.
- Le gf2i appelle à la désignation d'un organisme indépendant (tiers de confiance) auquel sera confié le détail des données utilisées par les systèmes d'IA génératives afin d'assurer la transparence.
- Le gf2i estime que les créations issues des IA ne peuvent bénéficier de la protection du droit d'auteur réservée à la seule création humaine.

#### **Introduction**

Le droit d'auteur (entendu ici comme incluant le droit d'auteur *stricto sensu*, les droits voisins et le droit *sui generis* des producteurs de bases de données) est constitué de l'ensemble des droits dont dispose le créateur d'une œuvre originale (et ses ayants droit) pour en maîtriser l'utilisation et la réutilisation (exploitation).

Les grands principes de ce corpus juridique ont plus de deux cents ans et ont fait preuve d'une grande souplesse en s'adaptant à de très nombreuses évolutions technologiques sans perdre leur A.D.N. Le **principe** fondamental de **l'autorisation préalable** à toute exploitation constitue le socle qui permet la définition de **modèles économiques adaptés**. Ce principe est assorti d'exceptions conditionnées. Ce cadre peut être schématisé comme une balance des intérêts, d'une part, des créateurs et producteurs et, d'autre part, des utilisateurs.

L'intelligence artificielle (IA) s'entend d'un ensemble de théories et de techniques visant à développer des programmes informatiques capables de résoudre des problèmes complexes sans fourniture explicite des algorithmes de résolution, c'est-à-dire de simuler l'intelligence humaine.

Les travaux d'Alan Turing sur la décidabilité annoncent l'intelligence artificielle, terme inventé en 1956 par le mathématicien John McCarthy, pionnier du domaine. Les assistants personnels intelligents (Siri, Alexa, etc.) apparus dans les années 2010 en sont les premières applications grand public.

C'est l'irruption et le développement phénoménal des usages d'intelligence artificielle générative dans la sphère publique depuis 2022 qui réouvrent la question de l'adéquation des dispositifs de protection des créateurs et producteurs de contenus.

En outre, la **convergence entre IA et moteur de recherche**, annoncée par certains opérateurs, fait craindre un véritable **changement de paradigme** pour les acteurs des industries culturelles et de la connaissance.

### ***L'IA, consommatrice compulsive de contenus protégés***

Pour fonctionner, les systèmes d'intelligence artificielle (SIA), en particulier générative (SIAG), doivent être entraînés sur d'importants volumes de données afin d'apprendre les relations entre celles-ci.

Ces données englobent du texte, de l'image (fixe ou animée), du son, du code, etc. Au sein de cet ensemble figurent des œuvres protégées par le droit d'auteur qui constituent des données structurées de grande qualité.

Les SIAG permettent d'accomplir des tâches en réponse aux requêtes (*prompts*) établis par les utilisateurs et aboutissent à la production de nouvelles données (textes, images, sons, etc.).

Les SIAG sont donc en contact avec le droit d'auteur sur toute la durée de leur cycle de fonctionnement : en amont, par l'utilisation d'œuvres protégées préexistantes, en aval, avec la production de nouveaux contenus.

En amont de la chaîne de traitements, les questions posées sont celles de la compatibilité des usages des différents intervenants de la chaîne de fonctionnement des SIAG avec les règles du droit d'auteur, celle du respect de ces règles.

### ***L'opt-out : condition nécessaire à la rémunération des ayants droit***

L'usage des œuvres n'est pas contesté par les plateformes opératrices (ce n'est pas leur ligne de défense dans les nombreux contentieux initiés aux USA par des auteurs et des éditeurs). La question centrale est celle de savoir si ces actes d'exploitation nécessitent ou non des

autorisations préalables, puis d'évaluer et de traiter les conséquences sur les modèles économiques des secteurs concernés.

En Europe, les articles 3 et 4 de la directive 2019/790 du 17 avril 2019 dite « DAMUN » (transposée à l'article L.122-5-3 du code de la propriété intellectuelle) ont prévu une exception<sup>1</sup> non compensée pour la « fouille de données » (« *text and data mining* » - TDM). On doit distinguer deux situations :

- La fouille effectuée à des fins de recherche scientifique par certaines personnes (organismes de recherche, bibliothèques, etc.) peut être réalisée librement sans autorisation des titulaires de droits ; dès lors qu'une entreprise dispose d'un accès privilégié à cette fouille, l'exception ne s'applique plus ;
- La fouille effectuée quelle que soit la finalité, y compris commerciale, est possible si l'accès au contenu est effectué de manière licite. En outre, les titulaires de droits ont la possibilité d'exercer un *opt-out* (un droit d'opposition) qui leur permet de neutraliser l'exception au droit exclusif d'exploitation.

En France, l'exercice de l'opt-out a été précisé par le décret n°2022-928 du 23 juin 2022. L'opposition du titulaire de droits n'a pas à être motivée et peut être exprimée par tout moyen. « Dans le cas de contenus mis à la disposition du public en ligne, cette opposition peut notamment être exprimée au moyen de procédés lisibles par machine, y compris des métadonnées, et par le recours à des conditions générales d'utilisation d'un site internet ou d'un service ».

De nombreuses oppositions ont ainsi été formulées, individuellement (conditions générales d'utilisation de sites internet de presse, par exemple) ou collectivement, notamment par des organismes de gestion collective, selon des modalités diverses. On soulignera que ces oppositions ont engendrés des coûts exclusivement supportés par les ayants droit à ce stade.

Le gf2i estime nécessaire un effort collectif pour rechercher et développer des **standards techniques** afin de donner toute son efficacité au dispositif. Ainsi, par exemple, TDM Reservation Protocol développé par le World Wide Web Consortium (W3C) existe d'ores et déjà et son usage est recommandé en France par le Syndicat national de l'Édition (SNE)<sup>2</sup>. Si la mise au point d'un standard unique semble peu probable, il est néanmoins souhaitable de parvenir à définir des **standards harmonisés** au niveau mondial.

Ce droit d'opposition peut être considéré comme un dispositif à vocation défensive destiné à s'opposer au scraping (ou web crawling) des données. Le gf2i considère qu'il constitue le moyen dont disposent les ayants droit pour parvenir à des **négociations avec les plateformes**,

---

<sup>1</sup> La question de savoir si, au moment de l'adoption de cette exception, le législateur européen l'envisageait comme couvrant l'IA générative nous semble aujourd'hui dépassée.

<sup>2</sup> [TDM Reservation Protocol \(TDMRep\) \(w3.org\)](https://www.w3.org/2022/07/tdm-reservation-protocol/)  
[Une clause-type pour s'opposer à la fouille de textes et de données par les intelligences artificielles - Syndicat national de l'édition \(sne.fr\)](https://www.sne.fr/actualites/une-clause-type-pour-s-opposer-a-la-fouille-de-textes-et-de-donnees-par-les-intelligences-artificielles)

afin de développer de **nouveaux marchés** pour une exploitation maîtrisée des œuvres par le développement de modèles économiques adaptés, sans entraver l'innovation.

L'effectivité du respect des droits par les plateformes d'IA repose donc fortement sur les **mesures préventives ou curatives** que prendront individuellement ou collectivement les ayants droits. Ces mesures consistent, par exemple, en des outils de surveillance, mesure et contrôle de l'activité des robots sur les sites internet ou encore de techniques de marquage ou de brouillage des données (œuvres) dont la charge financière devra, en premier lieu, être supportée par les ayants droit.

Bien entendu, ces coûts impacteront les modèles économiques devront être pris répercutés dans la construction des offres contractuelles d'accès aux contenus.

Ces nombreuses questions devront impérativement être traitées, notamment, dans le cadre de la mission du CSPLA confiée aux professeurs Alexandra BENSAMOUN et Joëlle FARCHY « relative à la rémunération des contenus culturels utilisés par les systèmes d'intelligence artificielle » chargée d'analyser les enjeux et d'examiner les mécanismes juridiques envisageables pour chaque secteur visant à garantir aux ayants droit l'effectivité de leurs droits lors de l'utilisation des œuvres par les fournisseurs d'IA.

#### ***La transparence sur l'entraînement des systèmes d'IA : nécessaire mais insuffisante***

Le projet de règlement européen établissant des règles harmonisées sur l'intelligence artificielle (RIA) ne revient pas sur la directive DAMUN. Il réaffirme la **nécessité de respect du droit d'auteur**<sup>3</sup> et instaure une **obligation de transparence sur les sources censée favoriser cet objectif**. Sa mise en œuvre doit, notamment, aider les ayants droit à disposer de moyens pour vérifier le respect de l'exercice de leur droit d'opposition. Cependant, les limites prévues par le texte<sup>4</sup> font douter de l'efficacité de cette obligation pour assurer le respect du droit d'auteur. La mission<sup>5</sup> du **CSPLA** dite « mission IA template » qui vient d'être confiée à la professeure Alexandra BENSAMOUN (elle doit permettre à la France de peser auprès de l'Office européen de l'IA en présentant des propositions pragmatiques) doit se donner pour objectif de contourner ces limites.

Cette **transparence** doit également porter sur les **périodes d'entraînement** des IA. On sait en effet, que les entraînements des SIAG sont déjà anciens pour certains, ce qui n'est pas sans conséquence sur le statut des utilisations d'œuvres :

---

<sup>3</sup> Considérant 106, article 53 1 (c).

<sup>4</sup> Considérant 107 à 109

<sup>5</sup> La mission a pour objet d'expertiser l'obligation de transparence faite aux plateformes d'IA générative par l'article 53 1 (d) du RIA qui consiste à établir et rendre disponible publiquement un « résumé suffisamment détaillé » des contenus utilisés pour l'entraînement de leurs modèles.

- Avant l'entrée en vigueur de la directive DAMUN (6 juin 2019), à défaut d'exception spécifique, le TDM d'œuvres protégées constituait une contrefaçon s'il n'a pas fait l'objet d'une autorisation ;
- Depuis l'entrée en vigueur de la directive DAMUN et jusqu'à la formulation d'un *opt-out* le TDM d'œuvres protégées bénéficie de l'exception ;
- Après la formulation d'un *opt-out*, le TDM d'œuvres protégées constitue une contrefaçon s'il n'a pas fait l'objet d'une autorisation.

Ce point n'est pas trivial, puisque certaines plateformes ont récemment indiqué de pas avoir conservé tous les jeux d'entraînement<sup>6</sup>. On observera que, paradoxalement, cette situation est conforme aux dispositions relatives à l'exception TDM.

On peut considérer que l'obligation de transparence prévue par le RIA permettra aux régulateurs et aux juridictions de s'assurer de la volonté de chaque opérateur de respecter le cadre juridique applicable et qu'il ne tire pas d'avantage concurrentiel avec les libertés prises avec les contraintes auxquelles d'autres acteurs se soumettent pour leur modèle.

La **transparence** favorisera également l'évaluation des qualités de chaque IA sur le marché en rendant possible la comparaison entre les données générées et celles ayant servi à l'entraînement des différents modèles, mais elle **ne saurait constituer une garantie pour les respect des droits des créateurs et producteurs de contenus protégés**.

### **Prévenir la concurrence déloyale des œuvres protégées par les productions artificielles**

S'agissant de l'aval de la chaîne, les productions des SIAG pourraient être confondues avec les œuvres protégées (de création humaine) et entrer en concurrence directe avec celles-ci, entraînant une destruction totale ou partielle de valeur.

Il résulte du cadre juridique international (et français) actuel que la protection du droit d'auteur ne peut bénéficier qu'à des auteurs personnes physiques<sup>7</sup>.

Les productions des SIAG ne peuvent donc bénéficier de la protection par le droit d'auteur. Le prompt – aussi précis soit-il – rédigé par l'utilisateur ne suffit pas à apporter un statut d'œuvre à la production générée. Par ailleurs, les dispositifs spécifiques de soutien, par exemple fiscaux, doivent être réservés aux œuvres protégées par le droit d'auteur afin de prévenir les phénomènes de concurrence déloyale qui résulterait de leur extension aux productions des SIAG.

Un dispositif de labellisation des productions synthétiques pourrait permettre d'informer le public et éviter toute confusion.

---

<sup>6</sup> [OpenAI a supprimé deux jeux de données utilisés pour entraîner illégalement GPT-3, ils contenaient des milliers de livres sous copyright et les employés qui avaient collecté ces données ont disparu \(developpez.com\)](#)

<sup>7</sup> Le cas des œuvres collectives n'y fait pas exception puisqu'il renvoie aux contributions individuelles.

Quant aux produits mixtes, il reviendra au juge d'apprécier la part de création humaine pour trancher les éventuels litiges.

### *Réviser le volet sanctions*

Les voies judiciaires à la disposition des ayants droit peuvent paraître inadaptées compte tenu de la vitesse de traitement des contentieux et de la vitesse de dissémination qui est celle du numérique. Il nous semble donc indispensable d'envisager d'examiner différentes pistes pour prévenir les usages non autorisés :

- Envisager de nouvelles compétences pour des autorités de régulation
- Renforcer les dispositifs de procédures rapides
- Étudier l'adoption de sanctions forfaitaires et provisionnelles

\* \* \*

Le gf2i estime que le cadre juridique existant offre des outils qui doivent être renforcés pour permettre, d'une part, le développement d'un marché des données respectueux du droit d'auteur et, d'autre part, la sanction des actes non autorisés, tout en favorisant le développement de nouveaux outils.

En outre, l'exercice de l'opt-out apparaît plus favorable à la **constitution d'un marché de licences** des contenus protégés – veillant à offrir des modèles adaptés aux différents secteurs et aux différentes utilisations – que porteur d'un « risque d'assèchement », notamment des contenus en langue française alors que ceux-ci ne constituent déjà qu'une très faible part des contenus présents dans les bases d'entraînements. Au contraire, le développement de modèles dans d'autres langues que l'anglais, notamment en français, ne pourra que profiter à la diversité culturelle.