

C O N T R O V **3** R S E

# RAPPORT À LA MISSION PARLEMENTAIRE PRADAL-RAMBAUD SUR L'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE

## Rapport intermédiaire sur la rémunération des auteurs et ayants-droits par les exploitants d'IA générative

sous la direction de Vincent Lorphelin  
fondateur du groupe de réflexion Controv3rse

AVEC LA CONTRIBUTION DE



# CONTACTS

Vincent Lorphelin, Fondateur, Controv3rse  
[contact@controv3rse.eu](mailto:contact@controv3rse.eu)

Jean-Michel Salomon, Président, Société des Auteurs de Jeux  
[jeanmichelsalomon@gmail.com](mailto:jeanmichelsalomon@gmail.com)

Jean-Philippe Clair, Directeur Marketing, Communication et Innovation, Keyrus  
[jean-philippe.clair@keyrus.com](mailto:jean-philippe.clair@keyrus.com)

Laurent Cervoni, Directeur de la Recherche et de l'Innovation, Talan  
[laurent.cervoni@talan.com](mailto:laurent.cervoni@talan.com)

Olivier Laborde, Leader Innovation et Transformation digitale, BPCE  
[olaborde@gmail.com](mailto:olaborde@gmail.com)

Jean Latger, Président Directeur Général, Oktal-SE  
[jean.latger@oktal-se.fr](mailto:jean.latger@oktal-se.fr)

Emmanuel Benazera, Président Directeur Général, Jolibrain  
[emmanuel.benazera@jolibrain.com](mailto:emmanuel.benazera@jolibrain.com)

Killian Vermersch, co-Fondateur et Directeur Général, Golem.ai  
[killian@golem.ai](mailto:killian@golem.ai)

Emmanuel Moyrand, co-Fondateur, France Meta  
[emmanuel@francemeta.org](mailto:emmanuel@francemeta.org)

Introduction	4
Synthèse	5
<b>1. Les auteurs protestent contre les IA génératives</b>	<b>7</b>
Des actions nombreuses et convergentes	7
Des revendications sur tous les chaînons	9
Une première manche emportée par les exploitants	10
<b>2. Le principe de la rémunération juste fait son chemin</b>	<b>11</b>
La solution la plus simple : utiliser les clés de répartition actuelles	11
Le principe de la rémunération juste	11
Quelques rares exploitants d'IAG ont déjà adopté ce principe	12
L'adoption de ce principe va probablement bénéficier d'enjeux d'images	12
<b>3. La mise en oeuvre de la rémunération juste n'est pas encore claire</b>	<b>13</b>
Des divisions latentes entre les représentants des auteurs	13
Un manque de clarté sur les objectifs poursuivis	15
Un manque de clarté sur la nécessité du degré de transparence des exploitants d'IAG	15
Une émergence encore peu lisible des modèles pionniers	15
Un risque de blocage du débat public	16
Les voies de déblocage du débat public	17
<b>4. La rémunération des auteurs doit faire l'objet d'un taux global</b>	<b>18</b>
Le pétrole du XXIème siècle, une métaphore utile	18
Un taux global plutôt que négocié par secteur	18
Une courte fenêtre d'opportunité	19
<b>5. La complexité intoxicque le débat public</b>	<b>20</b>
La spirale vertigineuse de la guerre des intelligences	20
La complexité, un point de vue obsolète	21
Vers une représentation simplificatrice	21
Le langage courant, véhicule de la pensée collective	22
<b>6. La « propriété liquide » propose un nouveau rapport de la société à la technique</b>	<b>23</b>
Le concept de « propriété liquide »	23
Une lecture simplifiée des enjeux	24
<b>Contributions</b>	<b>26</b>
Keyrus : « <i>La mise en oeuvre d'un modèle de rémunération équitable et transparent nécessite une collaboration étroite des parties prenantes</i> »	27
Talan : « <i>La piste de la transparence des concepteurs d'IAG nous apparaît comme la solution la plus simple</i> »	40
SAJE : « <i>Le Juge doit pouvoir accéder à la liste des « prompts » des utilisateurs, que ceux-ci revendiquent une paternité ou qu'ils soient suspectés de violer le droit d'auteur</i> »	51
BPCE : « <i>Nous allons voir apparaître de savantes formules mathématiques pour créditer et rémunérer les auteurs</i> »	52
Oktal-SE : « <i>L'idée serait de définir une métrique afin de qualifier, en fonction des données source, le curseur de plagiat</i> »	55
Jolibrain : « <i>Des points de mesure existent déjà, les concepts d'œuvre Essentielle et de Corpus National peuvent être considérés</i> »	58
Golem.ai : « <i>Des modèles alternatifs à l'IA générative doivent être privilégiés pour maintenir une traçabilité totale</i> »	64
France Meta : « <i>La rémunération équitable dans le domaine de l'IA générative ne concerne pas seulement les auteurs individuels, mais touche également à la santé globale de l'industrie de l'IA et à l'équité au sein de la société</i> »	65
<b>Bibliographie</b>	<b>68</b>



# INTRODUCTION

La création de la Mission d'information sur l'intelligence artificielle générative, sous la direction des rapporteurs Philippe Pradal et Stéphane Rambaud, a été annoncée le 30 mai 2023 par la Présidente de l'Assemblée Nationale Yaël Braun-Pivet : *« je pense avec effroi à cette pensée prophétique de Jean-Jacques Rousseau : 'Plus il y aura de machines pour faire le travail des Hommes, plus il y aura des Hommes qui ne seront que des machines.' Le doute et le questionnement, c'est donc à nous de les apporter ce soir. Tel est bien le but de 'l'Assemblée des idées' qui en est à sa cinquième édition. Telle est aussi la fonction de l'Assemblée nationale, qui se nourrit des travaux des chercheurs et des experts pour légiférer de la manière la plus pertinente possible»*.<sup>1</sup> La mission d'information sur l'intelligence artificielle générative analyse les défis qu'elle pose, en particulier concernant la protection des données personnelles et l'utilisation du contenu produit. Cette protection par un droit de propriété, souvent débattu<sup>2</sup>, rejoint celui des contenus d'entraînement des IA.

Jean-Noël Barrot, Ministre délégué chargé de la transition numérique et des télécommunications, s'était déjà prononcé sur la question de la rémunération des auteurs : *« Dans les mois et les années*

*qui viennent il nous faudra trouver une manière de faire en sorte que celles et ceux dont les contenus sont utilisés par les IA pour générer des contenus nouveaux puissent être dûment rétribués »*.<sup>3</sup>

Vincent Lorphelin anime les groupes de réflexion indépendants Controv3rse et Institut de l'Économie, et publie ses chroniques dans Le Monde et Les Echos. Auditionné par cette Mission, il souhaite apporter l'éclairage de ces groupes sur les revendications des auteurs (journalistes, écrivains, artistes, photographes, illustrateurs, programmeurs,...).

Ce rapport met en perspective technique, économique et sociale la rémunération des auteurs lorsque leurs créations sont utilisées pour entraîner ces IAG. Les aspects juridiques, qui relèvent des droits américain comme européen compte-tenu de leur convergence sur ce thème<sup>4</sup>, sont évoqués pour la seule contextualisation de notre analyse.

# SYNTHÈSE

Musiciens, photographes, scénaristes, écrivains, journalistes et programmeurs s'insurgent contre ChatGPT et autres intelligences artificielles génératives (IAG), qui s'entraînent avec leurs œuvres sans leur autorisation et sans les rémunérer. Pétitions, grèves et procès fleurissent.

Le camp des éditeurs d'IAG, à l'inverse, questionne la légitimité des auteurs à revendiquer des droits si les œuvres artificielles ne ressemblent pas aux œuvres initiales. On invoque la liberté d'apprentissage, d'inspiration et de style. On rappelle opportunément que la Cour suprême des États-Unis avait décidé en 2016 que la numérisation de livres par Google constituait un « usage équitable » et apportait « des avantages publics significatifs », ce qui justifiait de ne pas rémunérer les auteurs.

À Washington, le Sénat et le Copyright Office se sont saisis du problème, mais l'histoire est presque écrite d'avance. Le grand public préférera la magie des IAG gratuites aux protestations des ayants droit, comme aux premiers temps des échanges de musique en ligne, malgré les remontrances du gendarme Hadopi. On évoquera les débuts de la photographie, lorsque Charles Baudelaire s'insurgeait vainement contre son utilisation artistique. On arguera de la compétition avec les Chinois pour que l'innovation l'emporte sur les revendications corporatistes. Quelques concessions cosmétiques, comme la possibilité laissée aux auteurs d'interdire aux IAG d'utiliser leurs œuvres, achèveront le retour à la normale.



Pourtant ce scénario nous emmène tout droit vers la sous-culture. Les plateformes musicales sont déjà submergées par les morceaux artificiels. Les grandes bases de données d'images sont alimentées au rythme de 3 images artificielles pour 2 images d'auteurs. Comme 90 % du stock d'informations d'Internet a moins de 2 ans, la majorité des contenus sera bientôt artificielle. L'IAG se nourrira de ses propres contenus, dont la qualité sera triée par les traces d'intérêt laissées sur des plateformes comme TikTok. Les pertes d'originalité et de diversité seront à la hauteur des gains en productivité, et la culture artificielle suivra le triste chemin de la malbouffe.

Elon Musk, patron de Tesla et de Twitter, alerte sur une menace « existentielle » pour la survie de l'humanité. Au-delà du style outrancier qu'on lui connaît, il faut anticiper le jour où tout le monde pensera que la créativité n'est pas réservée aux humains. Le sentiment du déclasserement de l'humanité face aux machines rejoindra et amplifiera le pessimisme écologique.

Peut-on encore redresser cette tendance ? Pour commencer, il faut se souvenir que l'exception culturelle est une vraie invention française, et rappeler le mécanisme initié par Beaumarchais au moment de la Révolution pour rémunérer les auteurs : comme ils n'avaient pas les moyens d'identifier toutes les exploitations de leurs œuvres, il fallait des sociétés intermédiaires pour les détecter, collecter une part de la recette et la redistribuer. C'est ce qui permet aujourd'hui à un musicien de percevoir une rémunération lorsque son morceau est joué sur une radio à l'autre bout du monde. La France a conservé le leadership de cette gestion collec-

tive. Elle a encore été la première à initier puis transposer la directive européenne sur les droits d'auteur, et obtenir des plateformes la rémunération des extraits d'articles de presse. Sa puissance culturelle lui permet de battre le fer de l'IAG tant que le débat américain est chaud.

Une solution simple serait d'étendre le principe appliqué aux discothèques, qui consiste à prélever un pourcentage de leur chiffre d'affaires pour le reverser aux musiciens. Il suffirait de définir le taux équitable de la redevance à appliquer aux IAG et le reverser aux auteurs. L'économiste Ernst Fehr évalue par exemple à 14 % la valeur qu'apportent les agences de presse à Google Search, précurseur de l'IAG. Pour prendre un comparable plus établi, le taux de 15 % est celui que retiennent les pays producteurs de pétrole sur le prix du baril. On a suffisamment dit que les propriétés intellectuelles étaient le pétrole du XXI<sup>ème</sup> siècle pour ne pas s'en inspirer !

Pour être elle-même équitable, la répartition de cette collecte nécessite un mécanisme spécifique aux IAG. Une œuvre peut en effet être indispensable à une IAG alors que son exploitation directe est médiocre. Les euro-députés ont proposé que les IAG publient un « résumé détaillé de l'utilisation » des œuvres primaires. Il ne reste plus qu'à préciser quel contenu de ce résumé permettra d'authentifier ces œuvres, de mesurer leurs utilités relatives et de rémunérer proportionnellement les auteurs. Leurs propriétés deviendront aussi liquides que l'est devenu l'argent au XIX<sup>ème</sup> siècle, et la trajectoire de l'IAG se redressera vers un avenir plus désirable.<sup>5</sup>



## RECOMMANDATIONS

- 1** Établir la faisabilité d'algorithmes de rémunération, jugés équitables par les auteurs, en fonction de plusieurs hypothèses de niveaux de transparence des IAG.
- 2** Définir un taux global, applicable au chiffre d'affaires des exploitants d'IAG, pour rémunérer les auteurs et les ayant droits.
- 3** Proscrire des rapports officiels les mots « littérature », « art », « œuvre », « création » pour désigner des contenus synthétiques. Privilégier l'usage de la locution « propriété liquide » pour décrire la circulation économique des droits d'auteurs.



# LES AUTEURS PROTESTENT CONTRE LES IA GÉNÉRATIVES

## DES ACTIONS NOMBREUSES ET CONVERGENTES

Le tableau suivant liste les actions les plus visibles des auteurs et ayant-droits qui revendiquent la reconnaissance de leurs droits par les exploitants d'IA génératives.

### Tollé des auteurs et ayant-droits contre les IA génératives

DATE	DEMANDEUR	DÉFENDEUR	ŒUVRES	PROTESTATION
SEPT 2023	Sacem, SACD, Adagp, Adami + 70 organismes		Arts plastiques, musique, presse, audiovisuel, traduction, scénario, voix, théâtre, jeux, livres	Lettre <sup>6</sup>
	United Voice Artists		Acteurs, presse, traductions, auteurs, musique	Lettre <sup>7</sup>
	The Authors Guild George R.R Martin John Grisham	OpenAI	Livres	Procès <sup>8</sup>
	Michael Chabon David Henry Hwang	OpenAI	Livres, scénarios	Procès <sup>9</sup>
AOÛT 2023	New York Times	OpenAI	Presse	Pré-contentieux <sup>10</sup>
	Agence France-Presse European Press Photo Agency		Presse	Lettre <sup>11</sup>
	CNN Reuters Radio France	OpenAI	Presse	Blocage de GPTbot <sup>12</sup>
	Société des Auteurs et Compositeurs Dramatiques (SACD)		Audiovisuel et spectacle vivant	Déclaration <sup>13</sup>

DATE	DEMANDEUR	DÉFENDEUR	ŒUVRES	PROTESTATION
JUIL 2023	X (anciennement Twitter)	4 anonymes	Data	Procès <sup>14</sup>
	Cisac + 7 organismes		Musique	Lettre <sup>15</sup>
	Sarah Silverman Christopher Golden Richard Kadrey (Joseph Saveri)	Open AI	Livres	Procès <sup>16</sup>
	SAG-AFTRA Adami AEPO-ARTIS		Acteurs	Grève, déclaration <sup>17</sup>
	Class Action (Clarkson)	Google	Photos, vidéos et playlist personnelles	Procès <sup>18</sup>
JUIN 2023	Paul Tremblay, Mona Awad (Joseph Saveri)	Open AI	Livres	Procès <sup>19</sup>
	The Authors Guild	OpenAI Alphabet Meta Stability AI IBM	Livres, articles, essais, presse	Pétition <sup>20</sup>
MAI 2023	United Voice Artists		Voix off	Lettre <sup>21</sup>
	News Corps		Presse	Déclaration
	Writers Guild of America	Hollywood studios	Scénarios	Grève <sup>22</sup>
AVRIL 2023	Universal Music	Apple Spotify	Musique	Lettre
	Reddit		Conversation	Déclaration <sup>23</sup>
	European Writers' Council		Livres	Lettre <sup>24</sup>
	Asociación de Fotógrafos Profesionales de España		Photos	Déclaration <sup>25</sup>
MAR 2023	Human Artistry Campaign		Musique, images, livres	Pétition <sup>26</sup>
JAN 2023	Sarah Andersen, Kelly McKernan, Karla Ortiz	Stability AI Midjourney DeviantArt	Images	Procès <sup>27</sup>
	Getty Images	Stability AI	Images	Procès <sup>28</sup>
Nov 2022	Class action (Joseph Saveri)	GitHub Microsoft OpenAI	Programmes	Procès <sup>29</sup>

# DES REVENDICATIONS SUR TOUS LES CHÂÎONS

Le tableau suivant résume les principales revendications des auteurs et ayant-droits au regard des étapes du processus des IA génératives.

## Revendications des auteurs et processus des IAG

ŒUVRE PRIMAIRE	GRATTAGE (SCRAPING)	BASE DE DONNÉES	JEU DE DONNÉES D'ENTRAÎNEMENT (DATA SET)	APPRENTISSAGE	COMMANDE (PROMPT)	CONTENU SYNTHÉTIQUE (OUTPUT)
Autoriser	Collecte (opt in/out)			Entraînement	Noms d'auteurs et titres d'œuvres dans les prompts	
Identifier	Caractère personnel des data	Absence de reproduction des droits d'auteur ou copyright	Copie des œuvres	Data set et œuvres utilisés Droits d'auteurs et copyright		Relation entre œuvres et contenu synthétique
Lister	Œuvres collectées	Bases controversées	Origine du data set		Historique des prompts	
Contrôler			Déclarer les œuvres et data set		Provoquer un prompt pour générer une preuve d'utilisation	Étiqueter les contenus synthétiques
Rémunérer			Identifier les métriques pour l'algorithme	Critères de rémunération à dire d'expert, contribution au modèle	Noms d'auteurs et titres d'œuvres utilisés dans les prompts	Identifier les métriques de la valeur apportée par œuvre

Il n'y a pas d'inévitabilité technique, au sens informatique, pour la fourniture par les grands exploitants des informations correspondantes.

La Société des Auteurs de Jeux (SAJE) ajoute qu'il « devrait être possible pour le juge d'obtenir aisément de la part de l'exploitant de l'IAG la liste des 'prompts' des utilisateurs, que ceux-ci revendiquent une paternité ou qu'ils soient suspectés de violer le droit d'auteur »<sup>1</sup>.

On note cependant l'hétérogénéité des revendications entre groupes d'auteurs, ce qui multiplie les terrains de discussion avec les exploitants d'IAG. Ainsi Sam Altman, co-fondateur d'Open AI, observe que « les ayants droit veulent des choses différentes » : certains souhaitent que les logiciels comme ChatGPT ou Midjourney « publient un lien » vers leur œuvre, alors que d'autres veulent « qu'un fonds soit constitué et son argent distribué », et d'autres, encore, souhaitent « toucher une part des revenus si une IA crée une image dans le style d'un artiste ou une chanson dans le style des Beatles ».<sup>30</sup>

1. Voir la contribution de [SAJE](#).

## UNE PREMIÈRE MANCHE EMPORTÉE PAR LES EXPLOITANTS

Le procès intenté par les artistes Sarah Andersen, Kelly McKernan et Karla Ortiz contre Stability AI, Midjourney et Deviant Art, pour avoir abusé de leurs droits d'auteur, a été rapidement clos par un juge du tribunal fédéral de Californie. Il rejette l'idée qu'un contenu synthétique soit nécessairement une œuvre dérivée de celles qui ont servi à l'entraînement de son modèle. Il encourage en revanche les plaignants à démontrer des « similarités substantielles » entre contenus synthétiques et œuvres primaires<sup>31</sup>.

Mais les auteurs pensent disposer d'autres pistes pour faire progresser leurs revendications. « *L'IAG prend le travail de tout le monde et produit des œuvres qui pourraient le supplanter* »<sup>32</sup>, clament leurs avocats. Sur ce fondement, ils espèrent réfuter le caractère équitable<sup>33</sup> de l'utilisation de ce travail par les IAG.



# LE PRINCIPE DE LA RÉMUNÉRATION JUSTE FAIT SON CHEMIN

## LA SOLUTION LA PLUS SIMPLE : UTILISER LES CLÉS DE RÉPARTITION ACTUELLES

Pour ce qui concerne plus précisément la rémunération des auteurs, objet de notre rapport, il est largement admis que la collecte et la répartition de cette rémunération doit être faite par l'intermédiaire et selon le modèle des organismes de gestion collective (OGC).

Les clés de répartition actuelles des OGC pourraient-elles également être utilisées pour la rémunération provenant des IAG ? Le plus simple consisterait en effet à abonder le montant déjà collecté par ailleurs par les OGC. Si un auteur A perçoit aujourd'hui x % des montants distribués par un OGC dont il est membre, il percevrait toujours x % lorsque

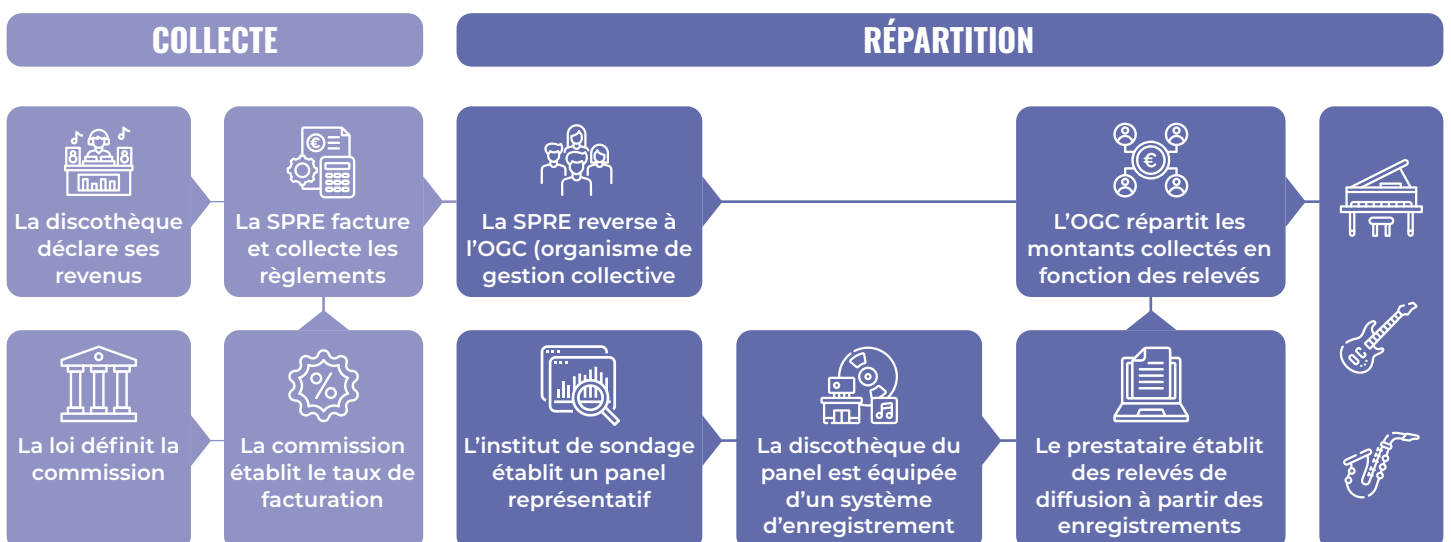
ces montants auront été augmentés par la redevance payée par les IAG.

C'est le schéma que suggèrent les débats en cours. Le Parlement Européen propose la mise à disposition du public d'un « résumé suffisamment détaillé de l'utilisation des données d'entraînement protégées par la législation sur le droit d'auteur »<sup>34</sup>. Si ce résumé ne contient pas de métriques des utilités relatives de ces données, il semble n'y avoir pas d'autre choix que d'utiliser les clés de répartition actuelles des OGC.

## LE PRINCIPE DE LA RÉMUNÉRATION JUSTE

Pourtant cette solution simple produirait inévitablement des désaccords. Les auteurs sont sensibles à l'équité de la répartition<sup>35</sup>. Pour la musique diffusée dans les lieux recevant du public, il aurait été facile en 1987 de recopier la clé de répartition sur les classements Club 40, Top club France

ou DJ top 100. Il a pourtant été décidé de constituer un échantillon de discothèques représentatif, qui établissent des relevés, pour que la clé de répartition soit au plus proche de la réalité.<sup>36</sup>



De même, pour la radio, la SACEM utilise un échantillon de radios considérées comme représentatives<sup>37</sup>, mais privilégie

quand c'est possible les relevés plutôt que les sondages<sup>38</sup> pour une rémunération plus juste<sup>(2)</sup>.

## QUELQUES RARES EXPLOITANTS D'IAG ONT DÉJÀ ADOPTÉ CE PRINCIPE<sup>(3)</sup>

Shutterstock est une des plus grandes banques d'images en ligne. Ses contributeurs reçoivent une part des redevances de licences aux exploitants d'IAG comme OpenAI (Dall-e) et LG (Exaone), qui sont des gros consommateurs d'images.<sup>39</sup>

Open AI a transposé ce modèle à la base de données des articles de presse de Associated Press.<sup>40</sup>

Adobe Firefly est une IA générative d'images. Les contributeurs dont le contenu a été utilisé par le premier modèle Firefly sont rémunérés.<sup>41</sup>

## L'ADOPTION DE CE PRINCIPE VA PROBABLEMENT BÉNÉFICIER D'ENJEUX D'IMAGES

Dès 2019, Apple, en se faisant le héraut de la confidentialité des données personnelles, a montré que la défense d'enjeux de société pouvait apporter un bénéfice d'image aux dépens de ses concurrents. Depuis, pour des raisons d'image, les Gafam ont tour à tour supprimé la reconnaissance faciale, accepté une sanction pour pratique commerciale trompeuse afin d'éviter l'étiquette d'obsolescence programmée,

renoncé à des commissions sur chiffre d'affaires d'applications au profit des petits développeurs, tout en organisant des campagnes de dénigrement de leurs concurrents.<sup>42</sup>

Le prochain terrain de cette bataille d'image pourrait être le principe de la rémunération juste, dans le sillage du modèle d'Apple Music depuis 2021.<sup>43</sup>

---

2. Voir la contribution de [France Meta](#).

3. Voir la contribution de [BPCE](#).








# LA MISE EN ŒUVRE DE LA RÉMUNÉRATION JUSTE N'EST PAS ENCORE CLAIRE

## DES DIVISIONS LATENTES ENTRE LES REPRÉSENTANTS DES AUTEURS

La mise en œuvre de la rémunération juste nécessite une règle de répartition fondée sur des indicateurs d'utilité relative des œuvres primaires, comme celle utilisée pour

les discothèques. Or ces indicateurs font l'objet d'une variété d'analyses, qu'illustre le tableau suivant.

## Modes de rémunération des auteurs par les IAG

EXEMPLES DE SOURCES	ATTRIBUTS DE L'ŒUVRE	VALEUR ÉCONOMIQUE DE L'ŒUVRE	VALEUR ÉCONOMIQUE DU JEU DE DONNÉES	RELATION DE L'ŒUVRE AU MODÈLE	UTILISATION DU MODÈLE	RELATION DE L'ŒUVRE AU CONTENU SYNT.	VALEUR ÉCONOMIQUE DU CONTENU SYNT.	PONDÉRATION DE L'ALGORITHME
								
Authors guild v. Open AI					Utilisation des noms, identité ou titres d'œuvres dans les prompts Durée d'utilisation du modèle	Utilisation de l'œuvre dans l'output		
Authors guild	Identité Nombre Longueur	Juste valeur Ventes						
Class action v. Google	Identité	Valeur marchande à dire d'expert			Durée d'utilisation du modèle			
Sacem, SACD + 70 organismes	Identité							
United Voice Artists + 12 organismes	Identité				Portée Objectif Durée d'utilisation			
Shutterstock	Volume des métadonnées		Vente d'ensemble de données					
Paul Tremblay v. Open AI							Profit	
Keyrus	Identité			Influence de l'œuvre sur la performance du modèle Fréquence ou importance d'utilisation de l'œuvre		Influence de l'œuvre sur l'output et sur sa qualité (pertinence, originalité)		Gouvernance des auteurs, utilisateurs et développeurs Dire d'expert
Talan	Identité volume				Données utilisées dans les prompts Durée d'utilisation			

Sources : Authors Guild v. open AI<sup>44</sup>, Authors Guild FAQs<sup>45</sup>, Class action v. Google<sup>46</sup>, Sacem, SACD, ADAGP, Adami + 70 organismes<sup>47</sup>, United Voice Artists + 12 organismes<sup>48</sup>, Shutterstock Contributors Fund (lancé en octobre 2022), p.m : Associated Press (accord avec OpenAI)<sup>49</sup>, contributions de Keyrus et Talan (voir plus loin).

Il y a un consensus sur le besoin d'identification des œuvres primaires.

Pourtant, au-delà, les représentants des auteurs se divisent. Certains semblent considérer cette identification comme suffisante pour l'exécution et la gouvernance des algo-

algorithmes de rémunération par les OGC des auteurs. D'autres suggèrent au contraire de compléter cette information par d'autres métriques (relation des œuvres aux modèles et aux outputs, degré d'utilisation du modèle, valeur économique des outputs).

## UN MANQUE DE CLARTÉ SUR LES OBJECTIFS POURSUIVIS

En dehors du principe de la rémunération juste, les variantes de revendications entre ces représentants révèle une hétérogénéité d'objectifs intermédiaires, par exemple :

- rémunérer les auteurs, dont les œuvres sont aspirées dans une base de données, au prorata de leurs revenus d'auteur connus
- rémunérer les auteurs, dont les œuvres sont utilisées pour l'entraînement d'un modèle, au prorata du volume des œuvres, mesuré en octets
- rémunérer les auteurs au prorata de la contribution individuelle de leurs œuvres aux profits générés par l'utilisation des IAG
- identifier les auteurs, dont les œuvres sont utilisées pour l'entraînement d'un modèle, pour négocier une rémunération en dehors d'autres indicateurs

## UN MANQUE DE CLARTÉ SUR LA NÉCESSITÉ DU DEGRÉ DE TRANSPARENCE DES EXPLOITANTS D'IAG

La probabilité d'aboutir à l'objectif d'une répartition jugée équitable par les auteurs augmente avec le degré de transparence des exploitants d'IAG, c'est-à-dire la quantité d'informations qu'ils fournissent.

Pourtant, imaginons demander à une IAG d'écrire un poème commençant par « Maître Corbeau ». On peut construire un indicateur de la ressemblance de ce poème synthétique avec la fable de la Fontaine, y compris sans transparence de l'IAG. Renouveler de nombreuses fois cette expérience, en variant les thèmes, permettrait peut-être de créer suffisamment d'indicateurs pour calculer une clé de répartition équitable.

Au besoin, d'autres informations relatives aux IAG sont largement disponibles sur le Web, souvent de manière gratuite : listes des œuvres accessibles au public, conditionnements (« prompt »), contenus synthétiques, code de modèle open

source, etc. Il est possible que leur analyse permette d'en extraire des indicateurs complémentaires<sup>4</sup>.

Rien ne prouve donc à ce stade le besoin de transparence de la part des exploitants d'IAG pour répartir une redevance ou une taxe, le cas échéant.

À l'inverse, rien ne prouve non plus que l'obtention de l'intégralité des indicateurs illustrés par le tableau précédent serait suffisante pour construire un algorithme de répartition acceptable par les ayant-droits. Seule l'adoption de modèles explicables et transparents apporterait cette garantie *a priori*<sup>5</sup>.

Dès lors, l'obtention de la liste détaillée des œuvres utilisées par les systèmes d'IAG et leurs sources<sup>50</sup> pourrait tout aussi bien se révéler suffisante, superflue ou insuffisante au regard de l'objectif.

## UNE ÉMERGENCE ENCORE PEU LISIBLE DES MODÈLES PIONNIERS

L'IAG est souvent décrite comme une boîte noire, puisqu'il est illusoire d'extraire d'un réseau de neurones<sup>6</sup> des règles de dépendance entre ses entrées et sorties. Cependant, la faisabilité technique d'une rémunération des auteurs des œuvres d'entraînement en fonction de leur utilité relative est déjà établie<sup>7</sup>. Quelques pionniers défrichent ce terrain :

### Shutterstock

Les contributeurs de la banque d'images reçoivent une part du prix de l'ensemble des contenus faisant l'objet d'une licence à un exploitant d'IAG. Cette part est proportionnelle au volume de ce contenu et de ses métadonnées : mots-clés, descriptions, géolocalisation, catégories, etc.<sup>51</sup>

4. Voir la contribution de [Jolibrain](#).

5. Voir la contribution de [Golem.ai](#).

6. Voir la contribution de [Jolibrain](#).

7. Voir les contributions de [Keyrus](#), [Talan](#), [BPCE](#), [Golem.ai](#), [Oktal-SE](#), [Jolibrain](#).

## Adobe Firefly

Les contributeurs dont le contenu a été utilisé dans la formation du jeu de données du premier modèle Firefly sont rémunérés en fonction du nombre total d'images, et des nombre et type de licences auxquels ces contenus ont contribué.<sup>52</sup>

En revanche, pour Adobe comme pour Shutterstock, l'assiette de la rémunération est une part du chiffre d'affaires de la banque et non de l'exploitant d'IAG. Il n'y a pas d'indicateurs sur les contenus synthétiques. Le pourcentage et l'algorithme ne sont pas publics.

## Nvidia Picasso et Bria

L'IAG Picasso NVIDIA utilise l'IA explicable Bria<sup>53</sup>, construite probablement sur un modèle de type symbolique<sup>(8)</sup>. Chaque fois qu'une image est générée, cette startup trace les visuels d'entraînement qui ont le plus contribué au contenu synthétique pour répartir les revenus entre les créateurs<sup>54</sup>.

L'assiette des revenus, le pourcentage reversé et l'algorithme ne sont pas publics.

## Ascendant Art

Cette startup, spécialisée dans la génération d'avatars, mise sur l'exigence de transparence par les auteurs. Elle leur rétrocède 25 % du prix de vente net d'un avatar, qu'ils se partagent au *pro rata* du poids de leurs œuvres dans cet avatar. Ce poids, c'est-à-dire l'utilité relative des œuvres

pour le contenu synthétique, est calculé par une autre intelligence artificielle<sup>55</sup>, à partir peut-être d'un curseur de plagiat<sup>(9)</sup>.

## Accord Associated Press / Open AI

Il s'agirait essentiellement d'un échange de marchandises : « l'accord prévoit que la licence OpenAI lui permet d'accéder aux archives de texte d'AP, tandis qu'AP tirera parti de la technologie et de l'expertise produit d'OpenAI ». De manière plus énigmatique, « AP soutient fermement un cadre qui garantira la protection de la propriété intellectuelle et la rémunération équitable des créateurs de contenu pour leur travail ».<sup>56</sup>

Ces comparables font émerger un modèle de rémunération des auteurs, à partir :

- de la liste des contenus d'entraînement
- d'indicateurs sur ces contenus : volume, mots-clés, descriptions, géolocalisation, catégories, etc.
- du chiffre d'affaires constituant l'assiette de rémunération
- d'un taux de rémunération des auteurs
- pour certains modèles, d'un indicateur de contribution de chaque œuvre d'entraînement à chaque contenu synthétique

Cependant, le manque de détails publiés sur ces comparables en gêne la lisibilité, en particulier celle du sentiment d'équité des auteurs.

# UN RISQUE DE BLOCAGE DU DÉBAT PUBLIC

Les euro-députés ont proposé que les IAG publient un « résumé détaillé de l'utilisation » des œuvres primaires. Il ne devait plus rester qu'à préciser quel contenu de ce résumé devait permettre d'authentifier ces œuvres, de mesurer leurs utilités relatives et de rémunérer proportionnellement les auteurs. Or le débat relatif à ce texte<sup>57</sup> risque de se bloquer, dans sa rédaction ou dans son application, entre deux groupes d'opinions opposées.

## Groupe pro-exploitants :

- il existe un risque que l'obtention des informations revendiquées par les auteurs, le cas échéant, soit vaine ou insuffisante au regard du principe recherché de rémunération juste

- un modèle open source offre suffisamment de transparence pour être interprétable par la communauté de ses utilisateurs<sup>58</sup>
- la légitimité de l'étendue des informations à publier par les exploitants d'IAG<sup>59</sup> manque de justifications
- la fourniture des informations revendiquées provoque un risque majeur pour les exploitants à cause de difficultés pratiques d'organisation, coût, secret des affaires et dynamique d'innovation<sup>60</sup>, hors impossibilité technique<sup>(10)</sup>.
- le législateur dispose de trop peu d'éléments pour apprécier le rapport entre légitimité des objectifs et difficulté pratique de la fourniture des informations.

8. Voir la contribution de [Colem.iq](#).

9. Voir les contributions de [Oktal-SE](#) et [BPCE](#).

10. Voir les contributions de [Keyrus](#) et [Talan](#).

### Groupe pro-auteurs :

- les modèles pionniers sont suffisamment matures pour être établis en tant que références

- le législateur a suffisamment d'éléments pour imposer la publication des indicateurs que ces modèles pionniers utilisent

## LES VOIES DE DÉBLOCAGE DU DÉBAT PUBLIC

Opposer innovation et culture, ou rater une opportunité de confirmer la puissance réglementaire de l'Europe lui serait certainement préjudiciable. Pour en limiter les risques :

### 1. Le groupe pro-exploitants devrait :

- clarifier les conditions dans lesquelles la fourniture des informations réclamées ne constituerait pas un risque majeur pour l'innovation et le développement des entreprises
- évaluer des algorithmes de rémunération juste des auteurs qui minimisent le besoin de ces informations

### 2. Le groupe pro-auteurs devrait :

- clarifier leurs objectifs intermédiaires,
- évaluer les solutions techniques pour réaliser une distribution équitable, en fonction de plusieurs hypothèses de niveau de transparence des IAG
- établir le niveau de transparence strictement nécessaire pour atteindre ces objectifs.

1

### RECOMMANDATION

Établir la faisabilité d'algorithmes de rémunération, jugés équitables par les auteurs, en fonction de plusieurs hypothèses de niveaux de transparence des IAG.



# LA RÉMUNÉRATION DES AUTEURS DOIT FAIRE L'OBJET D'UN TAUX GLOBAL

## LE PÉTROLE DU XXI<sup>ÈME</sup> SIÈCLE, UNE MÉTAPHORE UTILE

On dit depuis longtemps que les data et contenu d'Internet sont le pétrole du XXI<sup>ème</sup> siècle<sup>61</sup>. On parle de « matière première » pour désigner les inputs de l'intelligence artificielle<sup>62</sup>. Le chiffre d'affaires des GAFA est de 1 000 milliards de dollars<sup>63</sup>, c'est autant que le marché du pétrole<sup>64</sup>.

En poussant la comparaison en aval de l'extraction, ChatGPT correspondrait à l'invention de la bakélite en 1907, la première matière plastique industrialisable. L'IAG, comme l'industrie du plastique, génère un marché additionnel estimé à 1 000 milliards de dollars<sup>65</sup>.

Par contre, en amont de l'extraction, les différences sont frappantes, car les ordres de grandeur des redevances sont dans un rapport de 1 à 50.

Sur les 1 000 milliards de chiffre d'affaires des GAFA, seulement 3 milliards sont reversés en redevances<sup>66</sup>, soit 0,3 %, contre 150 milliards de redevances pour le pétrole, soit 15 %<sup>67</sup>.

Ce qui revient à dire que la propriété des auteurs est valorisée 50 fois moins que les gisements des pays pétroliers. On entend souvent que les contenus bruts n'ont pas de

valeur, que c'est leurs agrégation, croisement et contextualisation, c'est-à-dire leur exploitation, qui leur confèrent cette valeur<sup>68</sup>. Ceci justifie-t-il que l'économie de la donnée et de la connaissance profite à 99,7 % aux exploitants, non aux propriétaires ?

Les scénaristes et dialoguistes perçoivent 4 % du budget d'un film<sup>69</sup>. Les auteurs de livres perçoivent 8 % du prix du livre<sup>70</sup>. Les auteurs de théâtre perçoivent 12 % du prix des billets<sup>71</sup>. L'économiste Ernst Fehr<sup>72</sup> estime à 14 % la part équitable des revenus du moteur de recherche que Google devrait reverser aux éditeurs de presse<sup>73</sup>. Les auteurs, compositeurs et interprètes de musique perçoivent 27 % du chiffre d'affaires généré par le streaming<sup>74</sup>.

Jusqu'en 1960, le marché du pétrole était dominé par un oligopole de sept compagnies pétrolières qui contrôlait 85 % des réserves mondiales<sup>75</sup>. La création de l'OPEP a multiplié le prix du baril<sup>76</sup> par dix en dix ans<sup>77</sup>. Les revendications actuelles des auteurs sont-elles un signal de la revalorisation substantielle de leurs rémunérations ?

## UN TAUX GLOBAL PLUTÔT QUE NÉGOCIÉ PAR SECTEUR

Dès 2020, la Cour des Comptes constatait la dégradation de la situation économique et sociale des artistes-auteurs et l'érosion de leurs revenus, en dépit de l'augmentation générale de la valeur créée. Elle présentait un risque de surproduction et de destruction de valeur, en particulier avec l'apparition des œuvres engendrées par l'intelligence artificielle<sup>78</sup>.

Pour mieux reconnaître et répartir cette richesse, la directive européenne sur le droit d'auteur voulait garantir la part équitable des auteurs<sup>79</sup>.

La Cour des Comptes, observant que leur rémunération dépasse rarement 10 % des recettes d'exploitation, voulait rééquilibrer les rapports de force entre artistes-auteurs

(plasticiens, écrivains, scénaristes, compositeurs...) et acteurs de l'aval (producteurs, diffuseurs, éditeurs). Elle introduisait l'idée d'une instance interprofessionnelle, le « Conseil National des artistes auteurs » pour négocier des taux de référence par secteur<sup>80</sup>.

Le Député Guillaume Vuilletet propose une taxation des exploitants d'IAG, lorsque les œuvres primaires ne sont pas déterminées, à un taux fixé par le Conseil d'État<sup>81</sup>.

## UNE COURTE FENÊTRE D'OPPORTUNITÉ

La situation actuelle va probablement perdurer, au bénéfice des exploitants d'IAG et au dépens des auteurs.

Les recours collectifs peuvent prendre des années avant d'être jugés. Lorsque l'Authors Guild a intenté un procès contre Google en 2005, il a fallu sept ans pour que l'affaire soit entendue sur le fond. Le procès *In re Literary Works in Electronic Databases* a été lancé en 2001, mais les écrivains n'ont été indemnisés que dix-sept ans plus tard<sup>83</sup>.

Les internautes font massivement confiance aux IA génératives<sup>84</sup> et seulement 42 % d'entre eux pensent que les artistes devraient être rémunérés pour l'utilisation de leurs œuvres<sup>85</sup>. Lorsque Napster a permis le téléchargement de masse de la musique numérisée, le grand public a adopté

Pour limiter les phénomènes d'attrition, au bénéfice des contenus gratuits ou synthétiques, et décider d'un périmètre inexpugnable pour protéger la créativité, nous recommandons le principe d'un taux global<sup>82</sup>, fixé ou négocié avec les seuls exploitants d'IAG.

cet usage jusqu'à déstabiliser l'industrie, qui n'a finalement dû son salut qu'à Spotify et à des offres innovantes légales<sup>86</sup>.

Or le législateur est souvent soucieux que les contributions obligatoires soient indolores pour le consommateur<sup>87</sup>, ce qui rendra l'application d'une redevance ou d'une taxe d'autant plus difficile qu'elle sera tardive. Le fait que les offres d'IAG ne soient pas encore établies constitue donc une courte fenêtre d'opportunité avant que le marché ne s'organise et fixe les prix, ainsi que la répartition de la valeur, sur de mauvaises bases.

Ce taux devrait être fixé indépendamment de la définition des informations à obtenir des exploitants et du mode de répartition de la collecte.

**2 RECOMMANDATION**  
Définir un taux global, applicable au chiffre d'affaires des exploitants d'IAG, pour rémunérer les auteurs et les ayant droits.



# LA COMPLEXITÉ INTOXIQUE LE DÉBAT PUBLIC

## LA SPIRALE VERTIGINEUSE DE LA GUERRE DES INTELLIGENCES

L'écrivain Yuval Noah Harari pose la question : « *Que se passerait-il une fois qu'une intelligence non humaine deviendrait meilleure que l'Homme moyen pour raconter des histoires, composer des mélodies, dessiner des images et rédiger des lois et des écritures ?* » La réponse, selon lui, jette un nuage sombre sur l'avenir de la civilisation humaine<sup>88</sup>. L'IAG risque de devenir aussi dangereuse que « *les pandémies ou la guerre nucléaire* », poursuivent Elon Musk et Sam Altman, co-fondateurs d'OpenAI, Demis Hassabis, PDG de Google-DeepMind et des centaines de chercheurs de premier plan<sup>89</sup>. L'IA a bondi cette année de la quatorzième à la quatrième place parmi les « grands risques de demain »<sup>90</sup>. Elle constitue un risque systémique, au même titre que le réchauffement climatique est un risque écosystémique<sup>91</sup>, ce qui justifie pour le Premier Ministre britannique la création d'un « GIEC de l'IA »<sup>92</sup>.

Pour contrer Terminator<sup>93</sup> Elon Musk propose d'augmenter les capacités du cerveau grâce à des implants électroniques. Milliardaire justicier comme Batman<sup>94</sup>, son devoir serait d'intervenir en attendant que la police fasse son travail<sup>95</sup>.

Pour contrer le chômage de masse, Sam Altman préfère imaginer un « *capitalisme pour tous* » où la « *richesse phénoménale* » créée par l'IA serait redistribuée par l'intermédiaire d'une taxation des entreprises et des terrains fonciers<sup>96</sup>. Le recul de l'emploi serait pour sa part compensé par un « *revenu universel* »<sup>97</sup>.

Ce discours prospère sur le terrain de la techno-anxiété. La défiance de la société civile vis-à-vis des Gafa a été enclenchée par l'affaire Cambridge Analytica<sup>98</sup>. Elle s'est d'abord portée sur les données privées, les infos et les propos haineux<sup>99</sup>. Elle s'est ensuite étendue à la manipulation<sup>100</sup>, à la concentration des pouvoirs<sup>101</sup>, à la surveillance<sup>102</sup> et à l'affaiblissement de la démocratie<sup>103</sup>.

La fabrication massive de textes, images, discussions et sites Web personnalisés<sup>104</sup> alimentent la désinformation<sup>105</sup>, puis la sous-culture. Nourri par l'arrivée quotidienne de plus de 100 000 nouveaux titres, pour la plupart jamais écoutés, le catalogue de Deezer est passé de 90 à 200 millions de contenus ces deux dernières années<sup>106</sup>.

Le stock de contenu sur Internet va mécaniquement s'appauvrir en contenu primaire au bénéfice des contenus synthétiques<sup>107</sup>, produits en masse à bon marché. Le recyclage par d'autres IAG risque d'entretenir un cercle vicieux de perte de créativité, de diversité<sup>108</sup> et de personnalité.

Les auteurs apprivoisent toujours la technique, mais le risque nouveau introduit par l'IAG est celui d'une réduction de valeur plus rapide que la vitesse d'adaptation des auteurs. Le risque de perte artistique n'est pas directement lié à la technique, ce que plaidait Charles Baudelaire à propos de la photographie, mais à la spirale économique qu'elle provoque<sup>109</sup>. Goldman Sachs estime une destruction<sup>110</sup> de 2 % des emplois d'ici deux ans<sup>111</sup>.

La cognition a déjà été investie par la plateformes. Collecte et nettoyage de données en ligne, analyse de commentaires sur un produit ou une entreprise, modération de contenus, classement d'images et transcription d'enregistrements contribuent à alimenter et à renforcer l'intelligence artificielle, qui apprend à les exécuter elle-même<sup>112</sup>.

L'art algorithmique généré par ordinateur existe depuis les années 1960 mais les contenus synthétiques semblent maintenant créés par l'Homme<sup>113</sup>. Des photographies, tableaux ou musiques d'un rendu saisissant trompent aisément un œil ou une oreille humaine censée être exercée<sup>114</sup>. Les agences publicitaires et autres industries créatives adoptent

l'IAG<sup>115</sup>. Le créateur du groupe ABBA, aujourd'hui Président du premier réseau mondial des compositeurs de musique, reconnaît qu'il ne peut rien faire pour stopper l'IAG, ce qui pose d'« énormes questions existentielles »<sup>116</sup>.

Le cheval a disparu de l'économie alors qu'il en était le centre, on peut imaginer qu'il en sera de même pour l'humanité. Cette spirale est vertigineuse, car elle interroge la nature de l'Homme face à la machine. Existe-t-il une forteresse inexpugnable de l'humanité ?

## LA COMPLEXITÉ, UN POINT DE VUE OBSOLETE

Décrit dans ces termes, ce sujet est d'une complexité submergeante. Dès lors, les experts de la complexité en font un domaine réservé au nom de leur technicité, confisquant le débat et l'appropriation démocratique<sup>117</sup>. Elon Musk s'autorise même à tacler Mark Zuckerberg<sup>118</sup> et Bill Gates pour leur connaissance « limitée » de l'IA<sup>119</sup>.

Pourtant la complexité est le plus souvent le nom que nous donnons à des phénomènes dont nous ne parvenons pas à déceler la simple loi d'organisation<sup>120</sup>.

Avec l'hypothèse géocentrique (la Terre au centre), le soleil et la lune décrivent des arcs de cercles autour de la Terre,

mais Mars et les autres planètes décrivent des sortes de marguerites, avec des périodes de rétrogradations. Leurs équations complexes aboutissent à des prédictions approximatives de leurs mouvements. L'hypothèse héliocentrique (le soleil au centre) permet de calculer des trajectoires en ellipses. Non seulement les prédictions deviennent exactes, mais elles révèlent une loi simple et unique pour toutes les planètes : la loi de la gravitation universelle, qui a ouvert le champ immense de la physique moderne. Le géocentrisme n'est pas faux mais, générateur d'une complexité inutile, il est devenu un point de vue obsolète.

## VERS UNE REPRÉSENTATION SIMPLIFICATRICE

Pour analyser les enjeux de l'IAG et en démêler la complexité, il faut en chercher une nouvelle représentation. Nous proposons la piste suivante.

La question de savoir si une machine peut produire de l'art sera probablement débattue aussi longtemps que la philosophie de l'art. Pourtant, il existe des galeries d'art et des musées, donc en creux une définition sociale de l'art malgré l'absence de définition philosophique absolue. De même, il existe des protections légales des œuvres originales malgré l'absence de définition absolue d'une œuvre originale<sup>121</sup>.

Le législateur américain a ainsi rappelé à propos des tableaux « *Théâtre d'opéra spatial* » et « *A recent entrance to paradise* » que seul le produit d'une paternité humaine était susceptible de protection, et a conclu qu'un contenu synthétique n'était pas susceptible d'appropriation intellectuelle<sup>122</sup>. Cette décision

rejoint l'approche européenne qui estime que « les œuvres produites de manière autonome par des agents artificiels et des robots pourraient ne pas bénéficier de la protection par le droit d'auteur, afin de respecter le principe d'originalité, qui est lié à une personne physique, et étant donné que la notion de « création intellectuelle » porte sur la personnalité de l'auteur »<sup>123</sup>.



*Théâtre d'Opéra Spatial*

Autrement dit, s'il existe une forteresse inexpugnable du rôle économique de l'humanité dans les filières créatives, elle est fondée par une décision du législateur.

Il s'ensuit qu'il n'y a pas de compétition entre les intelligences, humaine et artificielle, pour la création de richesse intellectuelle. C'est ce qu'explique l'accord qui a mis fin à la grève des scénaristes de Hollywood, en stipulant que :

1. Les IAG ne sont ni des personnes, ni des écrivains. Leurs productions ne sont pas de la littérature.
2. Les travaux littéraires d'un écrivain ne changent pas de nature s'il utilise une IAG<sup>124</sup>. Ce qui revient à dire que l'IAG, comme un stylo ou un correcteur orthographique, n'est qu'un outil<sup>125</sup>.

## LE LANGAGE COURANT, VÉHICULE DE LA PENSÉE COLLECTIVE

Il paraît évident qu'un agriculteur qui utilise un tracteur reste un agriculteur et qu'un architecte qui utilise un logiciel 3D reste un architecte. Comment se fait-il qu'il fallait une grève puis un accord pour dire qu'un écrivain qui utilise une IAG reste un écrivain ?

La réponse est dans la langue, dont l'usage joue des tours à notre pensée collective.

Au départ, la vulgarisation des technologies nécessite un certain anthropomorphisme. Choisir la locution « intelligence artificielle » est certainement plus évocateur que « informatique statistique », mais elle place d'emblée la pensée dans des ornières dont cette dernière a ensuite un mal considérable à sortir.

Google poursuit la métaphore de l'intelligence artificielle pour affirmer que « les modèles d'IAG apprennent essentiellement à partir de toutes les informations disponibles. C'est un peu comme si un étudiant allait lire des livres dans une bibliothèque et apprenait ensuite à écrire et à lire »<sup>126</sup>.

Tandis que le registre de l'informatique statistique décrit l'entraînement des IAG comme la construction d'un modèle probabiliste.

La première locution guide l'intuition vers le caractère original des productions de l'IAG, la seconde vers une forme moderne de plagiat<sup>127</sup>. Cette dichotomie produit finalement un sentiment de malaise sociétal et de tromperie quand des productions de l'IA sont présentées comme le fruit d'une forme de créativité<sup>128</sup>. Ce sentiment signale un conflit entre le bon sens et nos grilles de représentation<sup>129</sup>, et invite à préciser notre vocabulaire.

Le Général de Gaulle, sachant que la langue est le véhicule de la pensée, s'agaçait des anglicismes utilisés dans l'armée, et demanda à son Ministre Pierre Messmer de les faire proscrire<sup>130</sup>.

De même, outre le mot « littérature », il faudrait proscrire des rapports officiels les mots « art », « œuvre » ou « création » lorsqu'ils désignent des contenus synthétiques.



# LA « PROPRIÉTÉ LIQUIDE » PROPOSE UN NOUVEAU RAPPORT DE LA SOCIÉTÉ À LA TECHNIQUE

## LE CONCEPT DE « PROPRIÉTÉ LIQUIDE »

Nous proposons le concept de « propriété liquide », pour guider l'intuition vers la forme de richesse que produisent les auteurs : insaisissable, irrigant l'ensemble de l'économie, spécifique à la création humaine, et non marchande tant qu'elle n'est pas exploitée. Il se réfère aussi à l'image de l'« argent liquide » diffusée par Emile Zola lors de la révolution industrielle<sup>131</sup>.

Ce concept prolonge les quatre grandes visions suivantes :

### L'économie de la créativité

A travers son Livre Vert<sup>132</sup>, l'Europe voulait « libérer le potentiel des industries culturelles et créatives ». A cette fin, elle déplorait que le numérique déplace la valeur économique vers le bout de la chaîne, ce qui altère la rémunération effective de la création. Elle appelait l'industrie à se doter de modèles innovants, qui commencent maintenant à apparaître.

Dans le cadre de l'application des droits voisins, Google négocie avec les éditeurs de presse les rémunérations des contenus (articles, photos, vidéos, etc.) affichés par son moteur de recherche.

Il leur communique à cette fin un « socle » d'informations comme le nombre d'impressions, le taux de clics sur le moteur de recherche et les données relatives aux revenus publicitaires.<sup>133</sup>

L'innovation des modèles n'est pas freinée par le manque d'informations disponibles. Les Big Tech disposent de

métriques sur l'utilité de chaque contenu pour chaque utilisateur<sup>134</sup>. Google Search et le fil d'actualité de Facebook<sup>135</sup> présentent les contenus dans l'ordre de leurs utilités, mesurées à partir des traces de leur utilisation (nombre de réactions, commentaires, partages) et de leur fiabilité (niveau d'expertise, caractère officiel, traçabilité<sup>136</sup>, etc.).

La finesse des métriques ouvre la perspective d'un internet dont tout utilisateur est reconnu en qualité de co-créateur lorsque sa production est révélatrice de sa personnalité, c'est-à-dire qu'elle fait naître un droit d'auteur.

La propriété liquide évoque la revalorisation des chaînons de la création grâce aux métriques d'utilité.

### L'économie du Web3

Les revendications actuelles des auteurs face aux exploitants de l'IAG rappellent celles de leurs prédécesseurs face aux plateformes du Web2. On se souvient que le Huffington Post avait commencé comme le Wikipedia des journalistes amateurs. Le nombre de clics permettait de mettre en avant les meilleurs articles. Lorsque le Huffington Post a été vendu 315 millions de dollars à AOL, les contributeurs se sont sentis lésés et ont fait un procès pour réclamer leur part. Ils ont été déboutés. Le Juge a en effet considéré que « les principes d'équité et de bonne conscience ne justifient pas que les plaignants reçoivent une partie du prix d'achat alors qu'ils ne s'attendaient pas à être payés, qu'ils ont accepté à plusieurs reprises le même marché et qu'ils se sont engagés dans l'arrangement les yeux grands ouverts ».<sup>137</sup>

Le modèle de plateforme a été souvent dénoncé pour conduire ses utilisateurs à travailler gratuitement<sup>138</sup> et à leur laisser la propriété des contenus.<sup>139</sup>

En réaction, le Web3 s'est explicitement construit autour de la blockchain en tant que contre-modèle, pour protéger les utilisateurs face à la concentration de pouvoir des plateformes.<sup>140</sup>

Les licences traditionnelles sont établies avec un exploitant et un produit identifiés, par exemple l'éditeur d'un livre. Les licences Web2 sont open source. Unilatérales, elles sont établies avec un exploitant ou pour un usage non identifiés *a priori*, ce qui leur confère une formidable capacité de diffusion. En contrepartie, elles sont souvent rentabilisées en dehors de cette licence par des ventes de services de formation ou de support technique. Les licences de la propriété liquide, avec ou sans blockchain, sont établies avec un exploitant ou pour un usage non identifiés *a priori*. Elles veulent combiner les avantages des deux premières : la viralité du Web2 et la rémunération des licences traditionnelles.

Ceci révèle un mouvement revendicatif puissant, qui pourrait s'exprimer de la manière suivante : « *Je suis prêt à laisser circuler ma propriété liquide librement à condition que, si quelqu'un en retire un bénéfice, il me verse une juste rémunération* ». <sup>141</sup>

### L'économie des micro-paiements

Dès ses débuts, Internet a été accompagné de solutions de micro-paiements pour tenter de remplacer le modèle payant du Minitel<sup>142</sup>. Face à l'abondance d'offres gratuites, le premier modèle viable a été celui de l'iPod d'Apple, pour acheter la musique à l'unité<sup>143</sup>. La baisse récente des coûts de transaction grâce à la blockchain fait espérer à Elon Musk<sup>144</sup>, patron de X, et à David Marcus, fondateur de la crypto-monnaie Libra<sup>145</sup>, une extension prochaine de ce modèle aux tweets et aux transferts d'argent.

## UNE LECTURE SIMPLIFIÉE DES ENJEUX

Avec le prisme de cette représentation simplifiée, peut-on avoir peur que l'IAG devienne « meilleure que l'Homme moyen pour raconter des histoires, composer des mélodies, dessiner des images et rédiger des lois et des écritures » ? L'ordinateur est déjà meilleur que l'homme pour la conduite d'un métro depuis 1983, au jeu d'échecs depuis 1997, au jeu de culture générale *Jeopardy* depuis 2011, au jeu de go depuis 2016 et pour le pilotage d'un avion depuis 2020. Toutes ces grandes avancées techniques, comme celles qui les ont précédées, ont toujours suscité un émoi qui semble

La CISAC veut mieux garantir la rémunération précise et équitable aux créateurs dans l'univers du streaming<sup>146</sup> grâce aux nouveaux identifiants des œuvres musicales (ISWC, GDSDX).

La blockchain et les NFT (*non fungible token*) ont la capacité intrinsèque d'assurer la traçabilité de la provenance et le paiement des royalties<sup>147</sup>.

Les micro-paiements, les nouveaux identifiants de contenus et la traçabilité dessinent une industrie de la créativité irriguée par des canaux de contenus, et régulée par les écluses des contrats intelligents<sup>148</sup>.

### L'économie du cerveau d'oeuvre<sup>149</sup>

Il ne s'agit pas de transhumanisme, mais d'une métaphore de l'Institut de l'économie pour transposer celle de la « main d'œuvre », née lors de la révolution industrielle précédente.

Une bonne part du travail d'un auteur est automatique. Un écrivain applique automatiquement les règles de l'orthographe et de la syntaxe, les références et citations viennent automatiquement à l'esprit du chercheur qui rédige un article, une part de l'activité mentale et manuelle d'un peintre ou d'un musicien est automatique. L'IAG peut assister ces auteurs pour prendre à sa charge ces automatismes. En revanche elle n'a pas de sensibilité ni d'intuition. Sans présence d'un auteur, ses productions manquent d'âme. « ChatGPT a lu des critiques de films, mais n'a jamais vu de film », dit un scénariste<sup>150</sup>. L'IAG permet aux auteurs de se concentrer sur « le propre de l'Homme ». Elle ouvre la perspective d'une symbiose entre intelligences artificielle et collective<sup>151</sup> pour produire une propriété liquide abondante et une nouvelle prospérité.

A la convergence de ces quatre grandes visions, la « propriété liquide » promet de devenir un carburant économique puissant.

un peu ridicule avec le recul du temps : l'imprimerie devait propager l'hérésie<sup>152</sup>, le chemin de fer devait provoquer des maladies<sup>153</sup>, l'électricité devait multiplier les incendies<sup>154</sup> et la photographie devait appauvrir le génie artistique<sup>155</sup>.

Les peurs que suscite l'innovation ne sont pas infondées, mais elles ne survivent pas à la domestication des outils, par leurs usage et réglementation, autrement dit quand l'innovation devient progrès.

Lorsque l'IAG sera domestiquée, la peur exprimée par Yuval Noah Harari ne disparaîtra pas, mais elle se focalisera sur l'innovation suivante et changera d'objet.

Que devient l'idée d'un revenu universel rendu obligatoire par le déclassement de l'humanité ? Lorsque le langage courant n'attribuera la créativité qu'à l'homme, l'idée de son déclassement se ringardisera et celle du revenu universel, qui circule depuis des siècles<sup>156</sup>, changera aussi d'objet.

Que devient le risque de recyclage infini des contenus synthétiques et de la sous-culture ? Lorsqu'il sera clair pour le grand public que l'IAG est un usage industriel de la propriété liquide, il paraîtra normal que cette exploitation implique des contributions directes à la production et à l'acquisition de droits sur les œuvres. L'esprit de la directive européenne « services de médias audiovisuels » a abouti à ce que Netflix contribue au financement de la production audiovisuelle<sup>157</sup>. Il sera transposé à l'IAG pour renouveler les contenus plutôt que les recycler, et promouvoir la diversité culturelle<sup>158</sup>.

Simplifier la représentation des enjeux de l'IAG n'en supprime pas les risques. Il permet en revanche de mieux les appréhender et doit contribuer à assainir le rapport de la société à la technique.

Résumons les simplifications que nous proposons pour représenter l'économie de l'IAG :

1. Les mots « littérature », « art », « œuvre » et « création » désignent des productions humaines
2. La richesse économique intellectuelle ne peut être le fruit que d'une production humaine
3. La nature de l'auteur ne change pas s'il utilise une IAG
4. La propriété liquide désigne la richesse intellectuelle en circulation dans l'économie numérique
5. L'IAG est un usage de la propriété liquide

### 3

## RECOMMANDATION

**Proscrire des rapports officiels les mots « littérature », « art », « œuvre », « création » pour désigner des contenus synthétiques. Privilégier l'usage de la locution « propriété liquide » pour décrire la circulation économique des droits d'auteurs**

# CONTRIBUTIONS

# KEYRUS : « LA MISE EN ŒUVRE D'UN MODÈLE DE RÉMUNÉRATION ÉQUITABLE ET TRANSPARENT NÉCESSITE UNE COLLABORATION ÉTROITE DES PARTIES PRENANTES »



Jean-Philippe Clair

Directeur Marketing, Communication et Innovation

## NOTE D'INTENTION

Quel éclairage apporter sur les moyens techniques (Web3, IA, filigrane, NFT, tokenisation, blockchain, traçabilité IA...) qui rendraient possibles la mise en œuvre des revendications des auteurs (journalistes, écrivains, artistes, photographes, illustrateurs, programmeurs...) dont les œuvres sont utilisées pour entraîner les IA génératives ?

Les technologies telles que Web3, IA, blockchain, NFT, tokenisation, etc., peuvent en effet jouer un rôle dans la mise en œuvre des revendications des auteurs dont les œuvres sont utilisées pour entraîner les IA génératives.

Après un premier partage de convictions sur le sujet, vous trouverez dans cette note d'intention comment ces technologies pourraient être appliquées.

## DES DÉFIS À LA FOIS INÉDITS ET ANCIENS POUR LA PROPRIÉTÉ INTELLECTUELLE DANS LE CONTEXTE DE L'INTELLIGENCE ARTIFICIELLE GÉNÉRATIVE

Lorsque vous utilisez une IA pour créer quelque chose de nouveau à partir de l'ensemble du savoir humain, il devient difficile de déterminer qui en est propriétaire et qui devrait être rémunéré. Ce problème est une préoccupation nouvelle, bien que les débats sur la propriété intellectuelle existent depuis des siècles.

Mais il ne faut pas sous-estimer la complexité du problème soulevé, et les défis à résoudre sont nombreux. Ainsi, on

pourrait s'appuyer sur plusieurs exemples tels que la création de chansons dans le style d'artistes célèbres à l'aide d'une IA. Dans ce cas, il est plus facile de comprendre les questions de propriété intellectuelle, mais l'utilisation d'IA génératives pour créer quelque chose dans le style de plusieurs artistes, ou même d'une époque pose des questions plus complexes. Qui devrait être rémunéré lorsque l'IA génère quelque chose dans le style des dix meilleures chansons de la dernière décennie, par exemple ?

Il existe également des problèmes similaires dans le domaine de l'art, où demander à une IA de créer une œuvre dans le style d'un artiste peut être vu comme du vol intellectuel par certaines personnes, tandis que d'autres estiment que cela n'est pas problématique, et que le fondement même de l'art est de se réapproprier et de réinterpréter les autres œuvres, le monde, le réel.

Une autre question importante concerne l'utilisation d'IA pour agréger et résumer des informations provenant de plusieurs sources. Les journaux, par exemple, peuvent considérer cela comme une utilisation de leurs contenus et demander une compensation. Cependant, les résumés, contrairement aux extraits, ne sont généralement pas soumis aux droits d'auteur.

L'IA générative ne reproduit pas directement les contenus sur lesquels elle a été formée, mais elle en déduit plutôt des modèles et des schémas. Ouvrons une parenthèse pour préciser ce point technique important.

En effet, la distinction entre «Training Data» (données d'entraînement) et «Model» (modèle) est fondamentale dans le domaine de l'apprentissage automatique, y compris dans le contexte de l'IA générative. Voici une explication approfondie de cette distinction :

### 1. Données d'entraînement (*Training Data*) :

- Les données d'entraînement sont le matériau brut utilisé pour enseigner à un modèle d'IA comment effectuer une tâche spécifique. Ces données sont essentielles pour former le modèle et lui permettre d'apprendre des schémas, des relations et des caractéristiques à partir des exemples.
- Dans le contexte de l'IA générative, les données d'entraînement peuvent inclure une vaste gamme d'informations, telles que du texte, des images, des vidéos, de la musique, etc. Ces données servent de base à l'IA pour générer du contenu similaire.
- Les données d'entraînement sont collectées à partir de sources diverses, telles que des livres, des articles, des enregistrements audio, des images sur le Web, etc. Plus les données d'entraînement sont riches et variées, plus le modèle d'IA sera capable de produire des résultats pertinents.
- Il est essentiel que les données d'entraînement soient de haute qualité et représentatives de la tâche que le modèle est censé accomplir. Des données de mauvaise qualité ou biaisées peuvent entraîner des résultats indésirables.

### 2. Modèle (*Model*) :

- Le modèle est la partie de l'IA qui est créée et ajustée lors de l'entraînement avec les données d'entraînement. Il s'agit d'un ensemble de paramètres, d'algorithmes et de poids qui permettent au modèle de comprendre et de générer des informations en fonction des connaissances acquises à partir des données d'entraînement.
- Une fois que le modèle a été formé, il peut être utilisé pour effectuer diverses tâches, telles que la génération de texte, d'images, de musique, la traduction automatique, la reconnaissance vocale, etc., en fonction de la nature de l'apprentissage et de la tâche spécifique pour laquelle il a été formé.
- Le modèle ne contient pas les données d'entraînement elles-mêmes, mais plutôt une représentation apprise de ces données. Il utilise cette représentation pour générer de nouvelles données ou effectuer des prédictions.
- Les modèles d'IA peuvent être réutilisés pour différentes tâches ou adaptés à des domaines spécifiques en réentraînant une partie du modèle avec de nouvelles données d'entraînement.

En résumé, les données d'entraînement sont les exemples bruts qui permettent au modèle d'IA d'apprendre, tandis que le modèle est l'entité qui intègre ces exemples pour générer des résultats. La distinction entre les deux est cruciale pour comprendre comment fonctionne l'apprentissage automatique et comment les modèles d'IA sont capables de générer du contenu basé sur leurs connaissances acquises.

L'IA fonctionne donc en agrégeant des quantités massives de données provenant de sources diverses, ce qui soulève des questions sur la rémunération équitable de ceux qui contribuent indirectement à ces données. La contribution d'une partie (personne, œuvre, corpus, ...) est-elle déterminante ? indispensable ? interchangeable ? nécessaire ? quantifiable ?

En fin de compte, il nous faut nous poser des questions sur la façon dont nous devrions aborder ces nouveaux défis en matière de propriété intellectuelle à l'ère de l'IA générative, notamment en ce qui concerne la rémunération, la réglementation et la perception de l'authenticité artistique dans un monde où les machines peuvent désormais « créer » de l'art. Pour ce faire, le concept de « propriété liquide », tel que proposé dans le présent rapport (cf. recommandation 3) s'impose donc naturellement pour décrire la circulation économique des droits d'auteurs pouvant s'appliquer dans ce cadre.

REVENDIGATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
<b>Système d'enregistrement / dataset</b>					
A 23. During training, the large language model copies each piece of text in the training dataset and extracts expressive information from it.	Identifier la copie de l'œuvre dans le dataset	KEY07	KEY08	KEY05 KEY06	KEY09 KEY11
A 29. BookCorpus, however, is a controversial dataset. It was assembled in 2015 by a team of AI researchers for the purpose of training language models. They copied the books from a Website called Smashwords.com that hosts unpublished novels that are available to readers at no cost. Those novels, however, are largely under copyright.	Identifier l'origine du training dataset. Établir une base des dataset controversés	KEY15 KEY16	KEY14		KEY12 KEY13
B 161. Courts recognize that Internet users have a property interest in their personal information and data. Plaintiffs' and Class Members' property rights in the personal data and information that they have generated, created, or provided through various online platforms thus includes the right to possess, use, profit from, sell, and exclude others from accessing or exploiting that information without consent or remuneration.	Identifier le caractère personnel des data	KEY16 KEY17			KEY16 KEY17
<b>Système d'enregistrement / LLM</b>					
B 21. Plaintiff N.G. reasonably expected that the information that he exchanged with these Websites would not be used by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff N.G. also did not consent to his private information contributed to Google products and services, including his Google searches, to be used to train the Products.	Identifier l'utilisation des données pour les compilations et entraînements	KEY15 KEY16	KEY14		KEY12 KEY13
B 82. Remarking on Web scraping practices like Defendants', Lapine highlighted the unique harm: "It's the digital equivalent of receiving stolen property... [my medical information] was scraped into this dataset... it's bad enough to have a photo leaked, but now it's part of a product."- More broadly, this "productization" of personal information means that all of the data about us scraped without permission from the full extent of our "digital footprints" is now fueling Bard's responses, to strangers around the world.	Identifier la collecte des données pour former les datasets. Identifier l'utilisation des datasets par l'IAG	KEY15 KEY16	KEY14		KEY12 KEY13

REVENDIGATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
<b>Système d'enregistrement / Output</b>					
A 5. Indeed, when ChatGPT is prompted, ChatGPT generates summaries of Plaintiffs' copyrighted works — something only possible if ChatGPT was trained on Plaintiffs' copyrighted works.	Provoquer un prompt qui génère une preuve d'utilisation d'une œuvre prédéfinie				
C 13. Copilot often simply reproduces code that can be traced back to open-source repositories or open-source licensees.	Identifier le code source dans l'output				
C 90. GitHub concedes that in ordinary use, Copilot will reproduce passages of code verbatim: "Our latest internal research shows that about 1% of the time, a suggestion [Output] may contain some code snippets longer than ~150 characters that matches" code from the training data.	Identifier les passages reproduits dans l'output et leur longueur		KEY19 KEY20 KEY21		
C 147. Defendants had access to but were not licensed by Plaintiffs nor the Class to create Derivative Works based upon the Licensed Materials. "Derivative Works" as used herein refers to Copilot's Output to the extent they are derived from Licensed Materials. The definition also includes the Copilot product itself, which is a Derivative Work based upon a large corpus of Licensed Materials.	Identifier la relation entre œuvre et œuvre dérivée. Celle-ci peut être un output ou le LLM lui-même		KEY22 KEY23	KEY22 KEY23 KEY24	
C 182. As a direct and proximate result of these material breaches by Defendants, Plaintiffs and the Class are entitled to an injunction requiring Defendants to comply with all the terms of any License governing use of code that was used to train Copilot, otherwise incorporated into Copilot, and/or reproduced as Output by Copilot.	Identifier l'œuvre et le copyright utilisée pour l'entraînement, incorporée dans l'IAG ou reproduite dans l'output	KEY15 KEY16	KEY14		
<b>Relevés</b>					
D. Create transparency obligations for AI developers to disclose what works they use to "train" their AI.	Déclarer les œuvres utilisées pour l'entraînement	KEY01 KEY02 KEY04 KEY10 KEY11	KEY03		
D. Require authors, publishers, platforms, and marketplaces to label AI-generated works and otherwise identify when a significant portion of a written work has been generated by AI.	Déclarer la nature générée par IA les œuvres ou portions significatives d'œuvres			KEY15 KEY16 KEY18	

REVENDEICATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
<b>Relevés</b>					
A 40. On information and belief, the reason ChatGPT can accurately summarize a certain copyrighted book is because that book was copied by OpenAI and ingested by the underlying OpenAI Language Model (either GPT-3.5 or GPT-4) as part of its training data.	Déclarer les dataset				
B 83. The remaining, substantial portion of Infiniset—a full 50%—is sourced from what Google vaguely terms as “public forums.” The company has declined to clarify the specifics of what constitutes these “public forums,” leaving users in the dark about the exact origins and nature of the data influencing half of the AI’s training.	Déclarer l’origine des données		KEY12 KEY13 KEY14 KEY16		
<b>Répartition</b>					
D. Compensate authors who wish to allow their works to be used in the “training” of generative AI (..) establish compensation for authors when their works are used in outputs, or when their names or identities or titles of their works are used in prompts.	Rémunérer l’auteur en fonction de l’utilisation de ses œuvres pour l’entraînement ou dans un output, ou lorsque leurs nom, identité ou titre est utilisé dans le prompt				KEY25
D. Once works have been used to “train” AI, they become part of the fabric of the AI and cannot be effectively removed. For the past “training” of what are called the “foundational” models—such as OpenAI’s GPT, Google’s LLaMDA and PaLM, and Meta’s LLaMA—the CMO will seek compensation on an annual basis going forward for as long as those foundational models are in use. “Training” AI on pre-existing works is not a one-time event; the works are constantly being used to continue to “train” the AI. As such, payment should continue as long as the AI model is in use.	Identifier la durée de vie du modèle AI				
D. Eligible writers who register with the CMO or a participating organization would receive a distribution based on algorithms that take into account the number of works published, the length of those works, and any available sales data.	Identifier des métriques (nombre, longueur et ventes des œuvres), définir le paramétrage et créer l’algorithme de répartition				

REVENDEICATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
<b>Répartition</b>					
AI companies should prevent use of creators' names, their writings (or portions thereof), or the titles of their work in prompts without the creators' express permission. And where writers permit this use, they need to be compensated.	Obtenir l'accord de l'auteur pour l'utilisation de ses nom, œuvres et titres d'œuvres dans les prompts. Rémunérer cette utilisation.	KEY15			
A 61. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits.	Rémunérer l'auteur en fonction du profit de chaque output				
C 156. The Licensed Materials add nearly all value to the Copilot product because the purpose of Copilot is to provide code and the source of that code is the Licensed Materials. Without the Licensed Materials, Copilot would not be functional.	Identifier les métriques de la valeur apportée par l'œuvre à l'output (ou un ensemble d'œuvres à un ensemble d'outputs)	KEY15 KEY25		KEY27	KEY25 KEY26 KEY27
C 180. Plaintiffs and the Class have suffered monetary damages as a result of Defendants' conduct. 165. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of the Licenses.	Identifier les métriques pour définir la clé de répartition des gains entre les œuvres.				KEY21 KEY24 KEY25 KEY26
C 183. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of contract.	Définir les métriques de la clé de répartition des gains entre les auteurs				
C 243. d) An award of statutory damages pursuant to 17 U.S.C. § 1203(b)(3) and 17 U.S.C. § 1203(c)(3), <sup>41</sup> or, in the alternative, an award of actual damages and any additional profits pursuant to 17 U.S.C. § 1203(c)(2) [The court shall award to the complaining party the actual damages suffered by the party as a result of the violation, and any profits of the violator that are attributable to the violation and are not taken into account in computing the actual damages, if the complaining party elects such damages at any time before final judgment is entered.]	Identifier les métriques du poids d'une œuvre dans la valeur d'un output (/d'un ensemble d'œuvres dans la valeur d'un ensemble d'outputs)				KEY19 KEY20 KEY21 KEY22 KEY24 KEY25 KEY26 KEY27

REVENDEICATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
<b>Répartition</b>					
B 86. In fact, the third largest site fueling the dataset is scribd.com, a subscription-based digital library with sixty (60) million e-books and audio books—that compensates authors using a revenue sharing model based on the number of reads their work gets. There is no indication Scribd consented to this mass misappropriation, and certainly the authors did not consent, nor were they compensated.	Rémunérer les auteurs sur un modèle de partage de revenus				
B 108. Instead of competing fairly, Defendants illegally copied the unique works of millions of creators to develop and “train” their AI technology, without consent, credit, or fair compensation. The Products’ ability to replicate the writing styles of specific authors, recreate the music and lyrics of specific musicians, duplicate the works of online content producers, as well as the ability to summarize and reproduce copyrighted materials, arises from the fact that these materials were copied by Defendants without authorization and injected into the underlying LLM as part of its training data.	Rémunérer chaque auteur en fonction de la contribution de son œuvre à l'entraînement de l'IAG				KEY19 KEY20 KEY21 KEY22 KEY24 KEY25 KEY26 KEY27 KEY28
B 190. To avoid the unjust enrichment of Defendants, this Court sitting in equity has the power to order a “data dividend” to consumers for as long as Bard and the Company’s other AI products generate revenue fueled on the misappropriated data. At the very least, Plaintiffs and the Classes should be personally and directly compensated for the fair market value of their contributions to the LLMs on which Bard was built, in an amount to be determined by expert testimony. Fundamental principles of property law demand such compensation, and everyday people reasonably support it.	Calculer la rémunération à partir de critères définis par experts. Rémunérer les auteurs en fonction de la valeur équitable de leurs contributions au LLM				KEY19 KEY20 KEY21 KEY22 KEY24 KEY25 KEY26 KEY27 KEY28 KEY29 KEY30 KEY31 KEY32 KEY33 KEY34

REVENDEICATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
<b>Autres revendications</b>					
D. The Authors Guild believes that it is inherently unfair to use and incorporate books, articles, and other copyrighted works in the fabric of AI technologies without the author's consent (..) or credit. (..) Because works were indiscriminately ingested for the existing foundational models, the Guild seeks the ability for the CMO to grant licenses on an opt-out basis—meaning the CMO can grant the license and obtain compensation on behalf of all professional writers, and those who do not wish to participate can opt out. Robust notice provisions to allow authors and other rights owners to opt out will have to be established.	Obtenir le consentement et citer l'auteur	KEY15			KEY15
A 41. At no point did ChatGPT reproduce any of the copyright management information Plaintiffs included with their published works.	Repérer l'absence de reproduction des droits d'auteur				
C 15. It violates the licenses that open-source programmers chose.	Identifier la licence du code source				
C 144. Without the authority of Plaintiffs and the Class, Defendants intentionally removed or altered CMI from the Licensed Materials after they were uploaded to one or more GitHub repositories.	Identifier la suppression des mentions de copyright				
C 170. Each of the Suggested Licenses requires at least (1) that attribution be given to the owner of the Licensed Materials used, (2) inclusion of a copyright notice for the Licensed Materials used, and (3) inclusion of the terms of the applicable Suggested License. When providing Output, Copilot does not comply with any of these terms.	Identifier l'œuvre, l'auteur et son copyright, inclure l'attribution et le copyright dans l'output				
B 87. Google's C-4 dataset also reflects the Company's deliberate receipt of stolen property to build and train Bard. The dataset contains data from "b-ok.org" a "notorious market for pirated e-books," as well as "[a]t least 27 other sites identified by the U.S. government as markets for piracy and counterfeits."	Identifier dans les datasets les données dont le site d'origine est notoirement frauduleux				
B 92. In February 2023, an official demonstration of Bard exposed the system's capacity to spread misinformation. In the demo, Bard was asked a question about the James Webb Space Telescope (JWST), in response to which it falsely asserted that JWST was the first to photograph exoplanets.	Filtrer les outputs de désinformation				

REVENDEICATION *	FONCTION TECHNIQUE	MOYEN TECHNIQUE			
		PAR L'AUTEUR	PAR LE DATASET	PAR LE LLM	PAR L'EXPLOITANT
B 93. After using the scraped personal data from millions of consumers to train the Products, Defendants did not stop there. Alarmingly, they continued to feed the Products by harnessing data gleaned from various of its own Google services, including Gmail and Google Search. Scraping of data from these platforms constitutes a pervasive and unconscionable invasion of users' personal spheres.	Filtrer les datasets provenant de Gmail et Google Search				
B 94. The average Gmail user had no idea that their private emails could be used for such purposes. Indeed, until relatively recently, generative AI products like Bard or Gemini were the province of science fiction. Now that some people are aware, they are frustrated that the Company does not allow any opportunity to opt-out of this collection of personal information as required by law.	Proposer un opt-out aux utilisateurs de Gmail				
B 139. Even if individuals could request that Bard remove their data, it is not possible to do so completely, because Defendants train Bard on individuals' inputs, personal information, and other users' data, which Defendants cannot reliably and fully extract from its trained AI systems any more than a person can "unlearn" the math they learned in sixth grade. Defendants have acknowledged this limitation explicitly, announcing last month that it is hosting a "machine unlearning challenge" for the Public to help figure it out since the inability to fully delete information can, in the words of Google, "raise privacy concerns."	Désapprendre à une IAG ou pallier un apprentissage abusif				
B 154. Bard's inclination to lie and spread misinformation also poses unique threats to all the authors and content creators whose works were stolen and embedded into the product. When Bard purports to regenerate the exact text of their works, sometimes it makes up portions. This can harm the author or creators' reputation by attributing to them things they never said or wrote. In all cases it interferes with the integrity of the work.	Vérifier l'attribution d'une citation à un auteur				

\* Chaque intitulé de revendication se réfère à une affaire et le numéro d'alignement des conclusions du plaignant :

A. *Paul Tremblay v. Open AI*

B. *class action v. Google*

C. *class action v. Github*

D. *Authors guild v. Open AI*

## POUR IDENTIFIER LES AUTEURS :

Il existe différentes technologies qui pourraient être utilisées seules ou combinées pour renforcer l'identification des auteurs et la protection de leurs œuvres dans le contexte de l'utilisation d'IA génératives. Voici quelques technologies pertinentes :

**KEY01 - Filigranes numériques (\*\*\*\*)** : Les auteurs pourraient intégrer des filigranes numériques dans leurs œuvres avant de les mettre à la disposition des systèmes d'IA. Ces filigranes pourraient être des codes uniques, des métadonnées spécifiques ou des modifications subtiles dans l'œuvre elle-même, ce qui permettrait de retracer l'origine de l'œuvre.

*Les filigranes numériques sont des techniques utilisées pour incorporer des informations discrètes et souvent imperceptibles dans des fichiers numériques tels que des images, des vidéos, des documents texte, etc. Ces informations ajoutées ne sont pas visibles à l'œil nu, mais elles peuvent être détectées et extraites à l'aide d'outils appropriés. Les filigranes numériques servent à diverses fins, notamment l'authentification, la traçabilité, la preuve de paternité et la protection contre la contrefaçon. Voici comment fonctionnent généralement les filigranes numériques :*

- **Incorporation** : Un filigrane numérique est généré en utilisant des algorithmes spécifiques qui produisent un modèle ou une séquence de données uniques et complexes. Ce filigrane est ensuite incorporé dans le fichier numérique cible, tel qu'une image ou un fichier audio. L'incorporation peut se faire à différents niveaux de l'œuvre, selon le type de filigrane utilisé.
- **Imperceptibilité** : L'objectif principal d'un filigrane numérique est de rester imperceptible pour les observateurs humains. Le filigrane est généralement conçu pour ne pas altérer significativement l'apparence ou la qualité du fichier original.
- **Robustesse** : Les filigranes numériques doivent être robustes contre diverses transformations ou altérations que le fichier pourrait subir, telles que la compression, la recadrage, la conversion de format, etc. Ils doivent être capables de résister à ces altérations tout en restant détectables.
- **Extraction** : Pour détecter et extraire un filigrane numérique, on utilise des logiciels spécialisés ou des algorithmes de détection. Ces outils analysent le fichier numérique à la recherche de motifs ou de signatures spécifiques qui correspondent au filigrane intégré.
- **Vérification** : Une fois le filigrane extrait, il peut être vérifié en le comparant au modèle original du filigrane. Si

*le filigrane correspond, cela peut confirmer l'authenticité de l'œuvre et prouver son origine.*

- Les filigranes numériques peuvent être basés sur différentes techniques, telles que la modification subtile des valeurs de pixels pour les images, l'ajout de données dans les plages inutilisées d'un fichier audio ou vidéo, ou même l'utilisation de méthodes de stéganographie pour cacher les informations au sein d'autres données. L'objectif est de créer une signature unique et difficile à contrefaire qui permet d'associer l'auteur à l'œuvre.
- Cependant, il est important de noter que l'efficacité des filigranes numériques peut dépendre de plusieurs facteurs, tels que la qualité de l'algorithme de filigrane, les capacités de détection, les transformations subies par l'œuvre et le niveau d'expertise des personnes cherchant à les contrefaire.

**KEY02 - NFT (Non-Fungible Tokens) (\*\*\*\*)** : Les auteurs pourraient créer des NFT pour leurs œuvres, enregistrant ainsi la propriété et l'authenticité de l'œuvre sur une blockchain. Les informations sur les auteurs pourraient être incluses dans les métadonnées du NFT.

**KEY03 - Registres décentralisés (\*\*\*\*)** : Une blockchain publique ou une technologie de registre décentralisé pourrait être utilisée pour enregistrer les informations d'identification des auteurs associées à leurs œuvres. Cela pourrait être un enregistrement immuable et transparent de la paternité des œuvres.

**KEY04 - Plateformes d'inscription (\*\*\*\*)** : Des plateformes spécifiques pourraient être mises en place où les auteurs peuvent enregistrer leurs œuvres avec des informations d'identification détaillées. Une fois que les œuvres sont enregistrées sur la plateforme, elles pourraient être liées aux systèmes d'IA génératives. Cela permettrait aux IA génératives d'accéder aux œuvres tout en conservant une trace de leur utilisation. La plateforme pourrait suivre l'utilisation des œuvres enregistrées par les systèmes d'IA génératives. Chaque fois qu'une œuvre est utilisée pour générer du contenu, la plateforme enregistre cette interaction. Si les œuvres enregistrées sont utilisées pour générer du contenu commercial ou d'une manière qui génère des revenus, les auteurs pourraient être rémunérés conformément aux accords établis sur la plateforme.

L'objectif principal des plateformes d'inscription est de fournir aux auteurs un moyen de prouver leur paternité, de suivre l'utilisation de leurs œuvres et de faciliter la rémunération équitable lorsque leurs œuvres contribuent à l'entraînement des IA génératives. Cela permettrait également aux utilisateurs

teurs de s'assurer qu'ils ont le droit d'utiliser ces œuvres de manière légale et respectueuse des droits d'auteur.

**KEY05 - Traitement de données (\*\*\*\*)** : L'analyse de texte et d'images pourrait être utilisée pour extraire automatiquement des informations sur les auteurs à partir des œuvres elles-mêmes, en se basant sur les éléments textuels ou visuels qui peuvent indiquer l'identité de l'auteur.

**KEY06 - Reconnaissance d'image et de texte (\*\*\*\*)** : Les technologies de reconnaissance d'image et de texte avancées pourraient être utilisées pour analyser les éléments visuels et textuels des œuvres afin d'identifier des caractéristiques uniques ou des informations liées aux auteurs.

**KEY07 - Systèmes d'empreintes digitales numériques (\*\*\*\*)** : Les empreintes digitales numériques sont des méthodes qui consistent à créer une signature unique pour chaque œuvre, basée sur des caractéristiques spécifiques de l'œuvre. Ces empreintes digitales pourraient être utilisées pour vérifier l'authenticité et l'origine des œuvres.

**KEY08 - Métadonnées avancées (\*\*\*\*)** : Enrichir les métadonnées des œuvres avec des informations détaillées sur

les auteurs, les dates de création, les licences d'utilisation et d'autres détails pertinents pourrait faciliter le suivi et l'identification.

**KEY09 - Stockage décentralisé (\*\*\*\*)** : Les technologies de stockage décentralisé, telles que InterPlanetary File System (IPFS) ou Filecoin, pourraient être utilisées pour stocker les œuvres et les métadonnées de manière sécurisée et immuable.

**KEY10 - Signature électronique (\*\*\*\*)** : Les auteurs pourraient utiliser des signatures électroniques pour attester de la paternité de leurs œuvres et de leur consentement à leur utilisation dans des systèmes d'IA génératives.

**KEY11 - Vérification d'identité (\*\*\*\*)** : Des solutions de vérification d'identité pourraient être intégrées dans les plateformes permettant aux auteurs de lier leurs œuvres à leur identité de manière vérifiable.

Il est important de noter que ces méthodes ne sont pas mutuellement exclusives et pourraient même être utilisées de manière combinée pour renforcer la preuve de paternité et l'authenticité des œuvres

## COMMENT TRACER L'UTILISATION DE LEURS ŒUVRES ?

Pour tracer l'utilisation des œuvres des auteurs dans le contexte de l'entraînement des IA génératives, diverses approches techniques peuvent être utilisées pour suivre et enregistrer les interactions entre les œuvres et les systèmes d'IA. Voici quelques méthodes possibles pour tracer l'utilisation des œuvres :

**KEY12 - Blockchain (\*\*\*\*)** : L'utilisation de la technologie de blockchain peut permettre de créer un registre immuable et transparent des interactions entre les œuvres et les systèmes d'IA. Chaque fois qu'une œuvre est utilisée pour l'entraînement d'une IA générative, une transaction peut être enregistrée sur la blockchain, indiquant les détails de l'interaction.

**KEY13 - Contrats intelligents (\*\*\*\*)** : Les contrats intelligents sur une blockchain pourraient être programmés pour enregistrer automatiquement les interactions entre les œuvres et les systèmes d'IA. Ces contrats pourraient également automatiser les processus de rémunération et de suivi en fonction des termes définis.

**KEY14 - Métadonnées embarquées (\*\*\*\*)** : Les œuvres pourraient être accompagnées de métadonnées spéciales qui enregistrent les détails de leur utilisation. Ces métadonnées pourraient être mis à jour chaque fois qu'une interaction a lieu.

**KEY15 - APIs et interfaces de suivi (\*\*\*)** : Les plateformes ou les systèmes d'IA génératives pourraient fournir des interfaces de suivi ou des API aux auteurs. Ces interfaces permettraient aux auteurs de vérifier quand et comment leurs œuvres ont été utilisées.

**KEY16 - Marquage des données (\*\*\*\*)** : Lorsque les œuvres sont utilisées pour entraîner des modèles d'IA, elles pourraient être marquées avec des informations de suivi spéciales. Ces marques pourraient être enregistrées dans les modèles générés et permettraient de retracer l'origine des données.

**KEY17 - Intégration avec les plateformes d'inscription (\*\*\*\*)** : Si les œuvres sont enregistrées sur des plateformes d'inscription, ces plateformes pourraient intégrer des fonctionnalités de suivi pour enregistrer les interactions et les utilisations.

En utilisant ces méthodes, les auteurs pourraient suivre avec précision comment leurs œuvres sont utilisées pour l'entraînement des IA génératives. Cela peut fournir une preuve solide de l'utilisation de leurs œuvres et soutenir les processus de rémunération équitable et de protection des droits d'auteur.

## COMMENT MESURER LE POIDS RELATIF DES ŒUVRES D'AUTEURS DANS LES ŒUVRES GÉNÉRÉES ?

Mesurer le poids relatif des œuvres d'auteurs dans les œuvres générées par les IA génératives peut être un défi complexe, mais il existe des approches techniques et analytiques qui peuvent aider à quantifier cette contribution. Voici quelques méthodes qui pourraient être utilisées pour estimer le poids relatif des œuvres d'auteurs dans les œuvres générées :

**KEY18 - Analyse de l'importance des données (\*\*)** : Les œuvres d'auteurs pourraient être évaluées en fonction de leur influence sur la performance globale du modèle d'IA générative. Des métriques de performance pourraient être utilisées pour mesurer comment les œuvres d'auteurs contribuent à la qualité, à la diversité ou à d'autres aspects des générations.

**KEY19 - Métriques d'influence (\*\*)** : Des métriques spécifiques pourraient être conçues pour évaluer l'impact des œuvres d'auteurs sur les sorties générées. Par exemple, on pourrait mesurer combien de phrases dans un texte généré ont été directement influencées par les œuvres d'auteurs.

**KEY20 - Analyse de perturbation (\*\*)** : Les œuvres d'auteurs pourraient être retirées sélectivement de l'ensemble d'entraînement, et l'effet sur la performance de l'IA générative pourrait être observé. Cela permettrait d'évaluer la diminution de qualité ou de pertinence dans les générations.

**KEY21 - Évaluation des contributions uniques (\*\*)** : En analysant la diversité et la variété des générations, on

pourrait identifier les parties des œuvres générées qui ne sont pas présentes dans les données de formation de l'IA, ce qui pourrait indiquer une contribution significative des œuvres d'auteurs.

**KEY22 - Attribution et désambiguïsation (\*\*)** : Des techniques d'attribution de contribution pourraient être utilisées pour estimer comment chaque œuvre a contribué à la formation de l'IA générative. Ces techniques visent à attribuer des parties spécifiques des générations à des œuvres particulières.

**KEY23 - Tests de sensibilité (\*\*)** : En introduisant des variations contrôlées dans les données d'entraînement, on pourrait observer comment ces variations influencent les générations. Cela aiderait à quantifier l'impact des œuvres d'auteurs par rapport aux autres données.

**KEY24 - Comparaisons avec d'autres données (\*\*\*)** : En utilisant des ensembles de données similaires mais sans les œuvres d'auteurs, on pourrait comparer les performances de l'IA générative pour évaluer l'effet spécifique des œuvres.

Il est important de noter que la mesure du poids relatif des œuvres d'auteurs dans les générations peut être complexe et subjective, car cela dépendra des objectifs spécifiques de l'utilisation de l'IA générative et des métriques choisies pour évaluer cette contribution.

## COMMENT RÉMUNÉRER CETTE UTILISATION ? COMMENT DÉFINIR LES CRITÈRES ET LES ALGORITHMES DE RÉMUNÉRATION ?

Rémunérer l'utilisation des œuvres d'auteurs dans l'entraînement des IA génératives est un défi complexe, mais il existe des approches qui peuvent être explorées pour définir les critères et les algorithmes de rémunération. Voici quelques idées sur la manière dont cela pourrait être abordé :

### DÉFINITION DES CRITÈRES DE RÉMUNÉRATION :

**KEY25 - Fréquence d'utilisation (\*\*\*\*)** : Les auteurs pourraient être rémunérés en fonction du nombre de fois que leurs œuvres sont utilisées pour l'entraînement des IA génératives. Les œuvres les plus utilisées pourraient recevoir une rémunération plus élevée.

**KEY26 - Importance de l'œuvre (\*\*)** : Les auteurs pourraient être rémunérés en fonction de l'importance ou de la pertinence de leurs œuvres dans les générations. Des critères d'importance pourraient être définis par la communauté ou les utilisateurs (ou le législateur !)

**KEY27 - Qualité des générations (\*\*)** : Si les œuvres d'auteurs contribuent à améliorer la qualité ou la pertinence des générations, les auteurs pourraient être rémunérés en fonction de l'impact positif sur la performance de l'IA générative.

**KEY28 - Innovation (\*\*)** : Les œuvres qui introduisent des concepts nouveaux ou originaux pourraient être rémunérées davantage, car elles pourraient apporter une valeur unique à l'IA générative.

### ALGORITHME DE RÉMUNÉRATION :

**KEY29 - Modèle basé sur les performances (\*\*)** : Un modèle de rémunération pourrait être conçu en fonction des métriques de performance de l'IA générative. Les œuvres d'auteurs contribuant positivement à la qualité ou à d'autres métriques pourraient être rémunérées en conséquence.

**KEY30 - Smart Contracts (\*\*\*\*)** : Des contrats intelligents sur une blockchain pourraient automatiser le processus de rémunération en fonction des critères définis. Chaque interaction entre une œuvre et une IA générative pourrait déclencher un paiement prédéfini.

**KEY31 - Pondération de critères (\*\*\*\*)** : Un système de pondération pourrait être utilisé pour combiner plusieurs critères, tels que la fréquence d'utilisation, l'importance et la qualité, afin de calculer le montant de la rémunération.

**KEY32 - Consensus communautaire (\*\*\*)** : Les critères de rémunération pourraient être définis collectivement par la communauté des auteurs, des utilisateurs et des développeurs (et le législateur ?) Les décisions pourraient être prises par le biais de processus de vote ou de gouvernance décentralisée.

**KEY33 - Évaluation qualitative (\*)** : Outre les métriques quantitatives, une évaluation qualitative par des experts ou des pairs pourrait également être utilisée pour évaluer la contribution des œuvres et déterminer leur rémunération.

## WEBLIOGRAPHIE SÉLECTIVE

Every Possible Melody Has Been Copyrighted, Stored On A Single Hard Drive, <https://www.hypebot.com/hypebot/2020/02/every-possible-melody-has-been-copyrighted-stored-on-a-single-hard-drive.html> : One of the reasons why it's so tough to defend a copyright lawsuit is because the court now considers a melody just a sequence of pitches, so Damien Riehl and Noah Rubin developed a program that recorded every possible melody (all 68.7 billion of them) via MIDI to a hard drive, but not for the reasons that you might think.

The Ai Crackdown Is Coming, Five ways for Washington to hold Silicon Valley accountable, By Matteo Wong, The Atlantic, <https://www.theatlantic.com/technology/archive/2023/08/ai-misinformation-scams-government-regulation/674946>

Generative AI and intellectual property, Benedict Evans, <https://www.benedict-evans.com/benedictevans/2023/8/27/generative-ai-ad-intellectual-property>

The State of AI Regulations in 2023, Holistic AI, <https://www.holisticai.com/papers/the-state-of-ai-regulations-in-2023>

Generative AI: Balancing Innovation and Intellectual Property Rights Protection, Dreyfus, <https://www.dreyfus.fr/en/2023/05/22/generative-ai-balancing-innovation-and-intellectual-property-rights-protection>

Copyright in front of artificial intelligence, Dreyfus, <https://www.dreyfus.fr/en/2021/12/13/work-created-by-an-artificial-intelligence-is-copyright-protection-possible>

Generative AI and the IP system - what does it all mean?, IP Australia, <https://www.ipaustralia.gov.au/news-and-community/news/2023/07/07/04/48/generative-ai-and-the-ip-system-what-does-it-all-mean>

Mais cette appréciation par les pairs est peut-être aussi la porte ouverte aux corporatismes et (petits) arrangements d'un autre âge...

**KEY34 - Révision périodique (\*\*)** : Les critères et les algorithmes de rémunération pourraient être revus périodiquement pour s'adapter aux évolutions de l'industrie, des technologies et des besoins des auteurs.

Il est important de noter que la mise en œuvre d'un modèle de rémunération équitable et transparent nécessite une collaboration étroite entre les auteurs, les utilisateurs, les développeurs de technologies, les juristes et d'autres parties prenantes. Les modèles choisis devront être éthiques, équilibrés et répondre aux objectifs de tous les acteurs impliqués.

### Critères d'évaluation de la faisabilité de chaque proposition :

- \* Impossibilité théorique ou solution non disponible
- \*\* Implémentation incertaine ou difficile
- \*\*\* Solution technique émergente (startup, test, POC, etc.)
- \*\*\*\* État de l'art

Generative AI's Intellectual Property Problem Heats Up > AIs producing art or inventions have to navigate a hostile legal landscape, and a consensus is far away, Rahul Rao, IEEE Spectrum, <https://spectrum.ieee.org/generative-ai-ip-problem>

Generative AI Has an Intellectual Property Problem, Gil Appel, Juliana Neelbauer, and David A. Schweidel, Harvard Business Review, <https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem>

Responsible Generative AI, Dataiku, [https://pages.dataiku.com/responsible-generative-ai?utm\\_id=14648494480--148950735297--665168671911--generative%20ai](https://pages.dataiku.com/responsible-generative-ai?utm_id=14648494480--148950735297--665168671911--generative%20ai)

Navigating intellectual property rights in the era of generative AI: The crucial role of educating judicial actors, UNESCO, <https://www.unesco.org/en/articles/navigating-intellectual-property-rights-era-generative-ai-crucial-role-educating-judicial-actors#:~:text=While%20the%20possibilities%20offered%20by,that%20resembles%20an%20existing%20creation>

Dialogue de l'OMPI sur la propriété intellectuelle et les technologies de pointe / WIPO Conversation on Intellectual Property and Frontier Technologies : The WIPO Conversation is an open, inclusive, multi-stakeholder forum intended to provide stakeholders with a leading, global setting to discuss the impact of frontier technologies on all IP rights and to bridge the existing information gap in this fast moving and complex field.

The eighth session of the WIPO Conversation will have a close look at the intersection of generative AI and IP and the challenges and opportunities it presents. It will address current issues, practical solutions, and future directions in the field.

Program : [https://www.wipo.int/meetings/en/doc\\_details.jsp?doc\\_id=614092](https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=614092)

# TALAN : « LA PISTE DE LA TRANSPARENCE DES CONCEPTEURS D'IAG NOUS APPARAÎT COMME LA SOLUTION LA PLUS SIMPLE »

Talan<sup>★</sup>

Laurent Cervoni

Directeur de la Recherche et de l'Innovation

## CONTEXTE

M. Vincent Lorphelin, auditionné dans le cadre de la Mission d'information sur l'intelligence artificielle générative, dirigée par ses rapporteurs MM. Philippe Pradal et Stéphane Rambaud, au sujet des moyens techniques qu'il serait possible de mettre en place pour

accéder aux revendications des auteurs quant à ces intelligences artificielles, a invité le Centre de Recherche et d'Innovation de Talan à partager son expertise sur ce sujet. Le présent document synthétise ces réflexions.

## DÉFINITION DES TERMES TECHNIQUES

Afin de prévenir toute confusion, nous listons ci-dessous les définitions retenues pour les différents termes techniques utilisés dans la suite du présent document.

- Une **intelligence artificielle générative**, abrégé en IAG, est un système informatique ayant la capacité de créer des données inédites (ou comportant au moins un certain degré de nouveauté) à partir d'une requête d'un utilisateur. Les données générées peuvent couvrir un large ensemble de modalités, incluant notamment le texte, le code informatique, l'image, la musique et la vidéo.
- Un **réseau de neurones** est un modèle général d'intelligence artificielle consistant en un certain nombre d'unités informatiques (appelées « neurones ») réparties en plusieurs couches, lesquelles sont reliées entre elles

par des fonctions mathématiques. Ces fonctions dépendent de valeurs numériques appelées **paramètres**, qui constituent le cœur du système : pour fonctionner, un réseau de neurones doit passer par une phase dite d'**entraînement**, au cours de laquelle une grande quantité de données existantes sont lues par le réseau et qui a pour but l'ajustement des paramètres. Ces ajustements sont réalisés automatiquement par le système lui-même, au fur et à mesure de la lecture des données d'entraînement. Les systèmes basés sur des réseaux de neurones constituent la plupart du temps l'état de l'art en intelligence artificielle.

- Un **grand modèle de langage**, abrégé en **LLM** (de l'anglais *Large Language Model*), est une instance d'IAG orientée vers la génération de données textuelles et composée d'un réseau de neurones caractérisé à la fois par

son nombre élevé de paramètres (au moins de l'ordre du milliard) et par la taille conséquente des données utilisées pour son entraînement. Les LLM représentent l'état de l'art en matière d'IAG textuelles.

Dans la suite du présent document, il sera supposé que les IAG considérées sont toujours basées sur des réseaux de neurones, sauf mention contraire explicite. On rappellera également que, dans ce cadre, les LLM constituent des cas particuliers d'IAG. On utilisera le terme « IAG déjà entraînée » pour désigner les IAG déjà existantes à l'heure actuelle et

## HYPOTHÈSES DU CADRE DE RÉFLEXION

Une IAG possède une forme de « connaissance » déterminée par les valeurs numériques prises par l'ensemble de ses paramètres au terme de l'entraînement, qui dépendent donc des données qui ont été fournies au système pour cette phase d'entraînement. Cependant, le consensus scientifique actuel établit qu'il est impossible de relier précisément ces valeurs à des fragments précis des données d'entraînement : l'ensemble des paramètres d'une IAG constitue une « boîte noire » dans laquelle on ne peut pas identifier la trace laissée par un jeu de données spécifique. On notera par ailleurs que pour la plupart des systèmes propriétaires, les paramètres ne sont pas accessibles au public, ou alors très partiellement lorsque l'entreprise a décidé d'en faire la publicité, justifiant d'autant plus la qualification de boîte noire.

Ce phénomène est d'autant plus important que le modèle est grand, ce qui s'applique tout particulièrement au cas des LLM et de la plupart des IAG mises à disposition du public par les très grandes plateformes. En conséquence, nous partons du principe qu'il est techniquement très difficile, si ce n'est impossible :

- (i) d'identifier quelles œuvres ont été utilisées au sein des données d'entraînement d'une IAG déjà entraînée sur l'unique base des paramètres de cette IAG ;
- (ii) de modifier les paramètres d'une IAG déjà entraînée dans le but d'effacer la contribution d'une œuvre utilisée lors de l'entraînement sans devoir reprendre cet entraînement à zéro.

Cet état de fait rend très incertain le processus de vérification qu'une œuvre donnée a été utilisée pour l'entraînement d'un IAG. Pour être très concret, ce n'est pas parce qu'un outil comme ChatGPT indique qu'il connaît l'œuvre de tel ou tel écrivain qu'il a réellement été entraîné sur les œuvres

dont la phase d'entraînement est achevée : ces IAG étant généralement proposées comme des produits ou des services par les grandes plateformes, il est très peu probable qu'elles subissent des modifications majeures de leurs paramètres par des phases complètes de réentraînement. Leurs concepteurs préféreront en effet travailler à de nouvelles versions d'IAG pour les remplacer, à l'image d'OpenAI qui a déployé GPT-4 pour faire suite à GPT-3.5 : s'il est probable que les données d'entraînement utilisées ont été les mêmes, les deux systèmes sont indépendants l'un de l'autre et n'ont pas été entraînés au même moment.

de cet écrivain. Il aura pu être entraîné par exemple sur des textes tiers qui mentionnent l'écrivain et son œuvre : l'outil étant entraîné pour créer des réponses crédibles à son interlocuteur, cette simple mention peut lui suffire à fournir une réponse contextuellement crédible. Il aura donc pu en l'espèce être entraîné à partir des œuvres de l'écrivain et/ou des textes d'analyse de l'œuvre de cet écrivain, voire aucune de ces deux solutions. Dans ce dernier cas, l'outil est juste en train d'inventer une réponse, ce qui illustre simplement la capacité générative de ces systèmes. Cette capacité générative est un comportement attendu des IAG qui résulte de leur dimension probabiliste, et leur permet notamment de générer des informations qui n'apparaissent pas dans les données d'entraînement : le cas est notamment flagrant lorsque des LLM inventent des dates dans la biographie de personnes, voire inventent l'existence même de certaines personnes lors des interactions avec l'utilisateur. Dans certains cas extrêmes, les LLM peuvent même générer des informations qui entrent en contradiction avec les données d'entraînement : ce phénomène, appelé « hallucination », retient de plus en plus l'attention des scientifiques qui étudient ces systèmes.

Considérant le caractère figé des IAG déjà entraînées, les difficultés techniques mentionnées ci-dessus impliquent qu'il n'existe pas de solution technique satisfaisante pour déterminer les œuvres utilisées pour l'entraînement de ces IAG, et que les progrès scientifiques n'apporteront probablement pas de méthodes efficaces pour contourner ces difficultés. En revanche, les IAG futures, pour lesquelles la phase d'entraînement n'a pas encore eu lieu, offrent davantage de possibilités pour le contrôle des données qui seront utilisées : les techniques mentionnées dans la suite de ce document concerneront donc principalement cette seconde catégorie d'IAG.

## ANALYSE

AFFAIRE	REVENDIGATION	FONCTION TECHNIQUE	
<b>Système d'enregistrement / dataset</b>			
Paul Tremblay v. OpenAI	23. During training, the large language model copies each piece of text in the training dataset and extracts expressive information from it.	Identifier la copie de l'œuvre dans le dataset	2
Paul Tremblay v. OpenAI	29. BookCorpus, however, is a controversial dataset. It was assembled in 2015 by a team of AI researchers for the purpose of training language models. They copied the books from a Website called Smashwords.com that hosts unpublished novels that are available to readers at no cost. Those novels, however, are largely under copyright.	Identifier l'origine du training dataset. Établir une base des dataset controversés	3, 6, 11, 22
Class Action v. Google	161. Courts recognize that Internet users have a property interest in their personal information and data. Plaintiffs' and Class Members' property rights in the personal data and information that they have generated, created, or provided through various online platforms thus includes the right to possess, use, profit from, sell, and exclude others from accessing or exploiting that information without consent or remuneration.	Identifier le caractère personnel des data	4, 7, 14
<b>Système d'enregistrement / LLM</b>			
Class Action v. Google	21. Plaintiff N.G. reasonably expected that the information that he exchanged with these Websites would not be used by any third-party looking to compile and use all his information and data for commercial purposes. Plaintiff N.G. also did not consent to his private information contributed to Google products and services, including his Google searches, to be used to train the Products	Identifier l'utilisation des données pour les compilation et entraînement	24
Class Action v. Google	82. Remarking on Web scraping practices like Defendants', Lapine highlighted the unique harm: "It's the digital equivalent of receiving stolen property. . . [my medical information] was scraped into this dataset. . . it's bad enough to have a photo leaked, but now it's part of a product." More broadly, this "productization" of personal information means that all of the data about us scraped without permission from the full extent of our "digital footprints" is now fueling Bard's responses, to strangers around the world.	Identifier la collecte des données pour former les datasets. Identifier l'utilisation des datasets par l'IAG	4, 7, 14, 24
<b>Système d'enregistrement / Output</b>			
Paul Tremblay v. OpenAI	5. Indeed, when ChatGPT is prompted, ChatGPT generates summaries of Plaintiffs' copyrighted works—something only possible if ChatGPT was trained on Plaintiffs' copyrighted works.	Provoquer un prompt qui génère une preuve d'utilisation d'une œuvre prédéfinie	28

AFFAIRE	REVENDICATION	FONCTION TECHNIQUE	
<b>Système d'enregistrement / Output</b>			
<b>Class action v. Github</b>	13. Copilot often simply reproduces code that can be traced back to open-source repositories or open-source licensees	Identifier le code source dans l'output	1, 8, 12
<b>Class action v. Github</b>	90. GitHub concedes that in ordinary use, Copilot will reproduce passages of code verbatim: "Our latest internal research shows that about 1% of the time, a suggestion [Output] may contain some code snippets longer than ~150 characters that matches" code from the training data.	Identifier les passages reproduits dans l'output et leur longueur	1, 12, 15
<b>Class action v. Github</b>	147. Defendants had access to but were not licensed by Plaintiffs nor the Class to create Derivative Works based upon the Licensed Materials. "Derivative Works" as used herein refers to Copilot's Output to the extent they are derived from Licensed Materials. The definition also includes the Copilot product itself, which is a Derivative Work based upon a large corpus of Licensed Materials.	Identifier la relation entre œuvre et œuvre dérivée. Celle-ci peut être un output ou le LLM lui-même	
<b>Class action v. Github</b>	182. As a direct and proximate result of these material breaches by Defendants, Plaintiffs and the Class are entitled to an injunction requiring Defendants to comply with all the terms of any License governing use of code that was used to train Copilot, otherwise incorporated into Copilot, and/or reproduced as Output by Copilot.	Identifier l'œuvre et le copyright utilisé pour l'entraînement, incorporée dans l'IAG ou reproduite dans l'output	
<b>Relevés</b>			
<b>Authors Guild v. Open AI</b>	Create transparency obligations for AI developers to disclose what works they use to "train" their AI	Déclarer les œuvres utilisées pour l'entraînement	
<b>Authors Guild v. Open AI</b>	Require authors, publishers, platforms, and marketplaces to label AI-generated works and otherwise identify when a significant portion of a written work has been generated by AI	Déclarer la nature générée par IA les œuvres ou portions significatives d'œuvres	16
<b>Paul Tremblay v. OpenAI</b>	40. On information and belief, the reason ChatGPT can accurately summarize a certain copyrighted book is because that book was copied by OpenAI and ingested by the underlying OpenAI Language Model (either GPT-3.5 or GPT-4) as part of its training data.	Déclarer les datasets	
<b>Class Action v. Google</b>	83. The remaining, substantial portion of Infiniset — a full 50 % — is sourced from what Google vaguely terms as "public forums." The company has declined to clarify the specifics of what constitutes these "public forums," leaving users in the dark about the exact origins and nature of the data influencing half of the AI's training.	Déclarer l'origine des données	

AFFAIRE	REVENDICATION	FONCTION TECHNIQUE	
<b>Répartition</b>			
<b>Authors Guild v. OpenAI</b>	Compensate authors who wish to allow their works to be used in the “training” of generative AI (..) establish compensation for authors when their works are used in outputs, or when their names or identities or titles of their works are used in prompts	Rémunérer l’auteur en fonction de l’utilisation de ses œuvres pour l’entraînement ou dans un output, ou lorsque leurs nom, identité ou titre est utilisé dans le prompt	21
<b>Authors Guild v. OpenAI</b>	Once works have been used to “train” AI, they become part of the fabric of the AI and cannot be effectively removed. For the past “training” of what are called the “foundational” models—such as OpenAI’s GPT, Google’s LLaMDA and PaLM, and Meta’s LLaMA—the CMO will seek compensation on an annual basis going forward for as long as those foundational models are in use. “Training” AI on pre-existing works is not a one-time event; the works are constantly being used to continue to “train” the AI. As such, payment should continue as long as the AI model is in use.	Identifier la durée de vie du modèle AI	
<b>Authors Guild v. OpenAI</b>	Eligible writers who register with the CMO or a participating organization would receive a distribution based on algorithms that take into account the number of works published, the length of those works, and any available sales data.	Identifier des métriques (nombre, longueur et ventes des œuvres), définir le paramétrage et créer l’algorithme de répartition	
<b>Authors Guild v. OpenAI</b>	AI companies should prevent use of creators’ names, their writings (or portions thereof), or the titles of their work in prompts without the creators’ express permission. And where writers permit this use, they need to be compensated.	Obtenir l’accord de l’auteur pour l’utilisation de ses nom, œuvres et titres d’œuvres dans les prompts. Rémunérer cette utilisation.	5, 21
<b>Paul Tremblay v. OpenAI</b>	61. Plaintiffs are entitled to statutory damages, actual damages, restitution of profits	Rémunérer l’auteur en fonction du profit de chaque output	
<b>Class action v. Github</b>	156. The Licensed Materials add nearly all value to the Copilot product because the purpose of Copilot is to provide code and the source of that code is the Licensed Materials. Without the Licensed Materials, Copilot would not be functional.	Identifier les métriques de la valeur apportée par l’œuvre à l’output (ou un ensemble d’œuvres à un ensemble d’outputs)	
<b>Class action v. Github</b>	180. Plaintiffs and the Class have suffered monetary damages as a result of Defendants’ conduct.  165. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of the Licenses.	Identifier les métriques pour définir la clé de répartition des gains entre les œuvres.	9

AFFAIRE	REVENDEICATION	FONCTION TECHNIQUE	
<b>Répartition</b>			
<b>Class action v. Github</b>	183. Plaintiffs and the Class are also entitled to recover as restitution from Defendants for any unjust enrichment, including gains, profits, and advantages that Defendants have obtained as a result of their breach of contract.	Définir les métriques de la clé de répartition des gains entre les auteurs	9
<b>Class action v. Github</b>	243. d) An award of statutory damages pursuant to 17 U.S.C. § 1203(b)(3) and 17 U.S.C. § 1203(c)(3), <sup>41</sup> or, in the alternative, an award of actual damages and any additional profits pursuant to 17 U.S.C. § 1203(c)(2) [The court shall award to the complaining party the actual damages suffered by the party as a result of the violation, and any profits of the violator that are attributable to the violation and are not taken into account in computing the actual damages, if the complaining party elects such damages at any time before final judgment is entered.]	Identifier les métriques du poids d'une œuvre dans la valeur d'un output (/d'un ensemble d'œuvres dans la valeur d'un ensemble d'ouputs)	
<b>Class Action v. Google</b>	86. In fact, the third largest site fueling the dataset is scribd.com, a subscription-based digital library with sixty (60) million e-books and audio books — that compensates authors using a revenue sharing model based on the number of reads their work gets. There is no indication Scribd consented to this mass misappropriation, and certainly the authors did not consent, nor were they compensated.	Rémunérer les auteurs sur un modèle de partage de revenus.	5, 9, 17
<b>Class Action v. Google</b>	108. Instead of competing fairly, Defendants illegally copied the unique works of millions of creators to develop and “train” their AI technology, without consent, credit, or fair compensation. The Products' ability to replicate the writing styles of specific authors, recreate the music and lyrics of specific musicians, duplicate the works of online content producers, as well as the ability to summarize and reproduce copyrighted materials, arises from the fact that these materials were copied by Defendants without authorization and injected into the underlying LLM as part of its training data.	Rémunérer chaque auteur en fonction de la contribution de son œuvre à l'entraînement de l'IAG	5, 9
<b>Class Action v. Google</b>	190. To avoid the unjust enrichment of Defendants, this Court sitting in equity has the power to order a “data dividend” to consumers for as long as Bard and the Company's other AI products generate revenue fueled on the misappropriated data. At the very least, Plaintiffs and the Classes should be personally and directly compensated for the fair market value of their contributions to the LLMs on which Bard was built, in an amount to be determined by expert testimony. Fundamental principles of property law demand such compensation, and everyday people reasonably support it.	Calculer la rémunération à partir de critères définis par experts. Rémunérer les auteurs en fonction de la valeur équitable de leurs contributions au LLM	13, 18

AFFAIRE	REVENDICATION	FONCTION TECHNIQUE	
<b>Autres revendications</b>			
<b>Authors Guild v. OpenAI</b>	The Authors Guild believes that it is inherently unfair to use and incorporate books, articles, and other copyrighted works in the fabric of AI technologies without the author's consent (..) or credit. (..) Because works were indiscriminately ingested for the existing foundational models, the Guild seeks the ability for the CMO to grant licenses on an opt-out basis — meaning the CMO can grant the license and obtain compensation on behalf of all professional writers, and those who do not wish to participate can opt out. Robust notice provisions to allow authors and other rights owners to opt out will have to be established.	Obtenir le consentement et citer l'auteur	15, 25
<b>Paul Tremblay v. OpenAI</b>	41. At no point did ChatGPT reproduce any of the copyright management information Plaintiffs included with their published works.	Repérer l'absence de reproduction des droits d'auteur	12, 19, 23
<b>Class action v. Github</b>	15. It violates the licenses that open-source programmers chose	Identifier la licence du code source	12, 19, 23
<b>Class action v. Github</b>	144. Without the authority of Plaintiffs and the Class, Defendants intentionally removed or altered CMI from the Licensed Materials after they were uploaded to one or more GitHub repositories.	Identifier la suppression des mentions de copyright	12, 19, 23
<b>Class action v. Github</b>	170. Each of the Suggested Licenses requires at least (1) that attribution be given to the owner of the Licensed Materials used, (2) inclusion of a copyright notice for the Licensed Materials used, and (3) inclusion of the terms of the applicable Suggested License. When providing Output, Copilot does not comply with any of these terms.	Identifier l'œuvre, l'auteur et son copyright, inclure l'attribution et le copyright dans l'output	15
<b>Class Action v. Google</b>	87. Google's C-4 dataset also reflects the Company's deliberate receipt of stolen property to build and train Bard. The dataset contains data from "b-ok.org" a "notorious market for pirated e-books," as well as "[a]t least 27 other sites identified by the U.S. government as markets for piracy and counterfeits."	Identifier dans les datasets les données dont le site d'origine est notoirement frauduleux	6, 22, 26
<b>Class Action v. Google</b>	92. In February 2023, an official demonstration of Bard exposed the system's capacity to spread misinformation. In the demo, Bard was asked a question about the James Webb Space Telescope (JWST), in response to which it falsely asserted that JWST was the first to photograph exoplanets.	Filtrer les outputs de désinformation	15, 27

AFFAIRE	REVENDICATION	FONCTION TECHNIQUE	
<b>Autres revendications</b>			
<b>Class Action v. Google</b>	93. After using the scraped personal data from millions of consumers to train the Products, Defendants did not stop there. Alarming, they continued to feed the Products by harnessing data gleaned from various of its own Google services, including Gmail and Google Search. Scraping of data from these platforms constitutes a pervasive and unconscionable invasion of users' personal spheres	Filtrer les datasets provenant de Gmail et Google Search	10, 20
<b>Class Action v. Google</b>	94. The average Gmail user had no idea that their private emails could be used for such purposes. Indeed, until relatively recently, generative AI products like Bard or Gemini were the province of science fiction. Now that some people are aware, they are frustrated that the Company does not allow any opportunity to opt-out of this collection of personal information as required by law.	Proposer un opt-out aux utilisateurs de Gmail	25
<b>Class Action v. Google</b>	139. Even if individuals could request that Bard remove their data, it is not possible to do so completely, because Defendants train Bard on individuals' inputs, personal information, and other users' data, which Defendants cannot reliably and fully extract from its trained AI systems any more than a person can "unlearn" the math they learned in sixth grade. Defendants have acknowledged this limitation explicitly, announcing last month that it is hosting a "machine unlearning challenge" for the Public to help figure it out since the inability to fully delete information can, in the words of Google, "raise privacy concerns."	Désapprendre à une IAG ou pallier un apprentissage abusif	
<b>Class Action v. Google</b>	154. Bard's inclination to lie and spread misinformation also poses unique threats to all the authors and content creators whose works were stolen and embedded into the product. When Bard purports to regenerate the exact text of their works, sometimes it makes up portions. This can harm the author or creators' reputation by attributing to them things they never said or wrote. In all cases it interferes with the integrity of the work.	Vérifier l'attribution d'une citation à un auteur	1, 12, 15

#### Commentaires du tableau :

1. L'utilisateur final de l'IAG pourrait, à l'aide d'une recherche Internet, vérifier l'exactitude d'une citation ou repérer des citations non sourcées.
2. Création d'empreintes numériques uniques pour chaque morceau de texte dans l'ensemble des données
3. Les auteurs de contenu devraient pouvoir associer, manuellement ou non, leur contenu au type de controverse auxquels ils sont sujets
4. L'auteur du dataset pourrait mettre en place des méthodes techniques pour identifier et signaler les informations personnelles au sein des données et documenter ces identifications
5. Les concepteurs devraient fournir une base de données signalant la provenance et le copyright de leur corpus

6. Le dataset devrait comprendre les empreintes de chaque source de données utilisées avec les zones extraites à une date donnée.
7. Le dataset pourrait inclure des métadonnées spécifiant quelles données sont considérées comme des données personnelles, ce qui aiderait les utilisateurs à les identifier
8. Le corpus devrait ajouter à chaque code source collecté l'objectif du code, la source et d'autres métadonnées nécessaires
9. Le corpus devrait contenir, pour chaque ressource utilisée, quelle proportion du dataset final cela représente afin de calculer le pourcentage de revenus de chaque auteur.
10. L'ensemble de données lui-même pourrait inclure des métadonnées spécifiant les provenances des données
11. Le LLM (ou un modèle externe) doit être entraîné pour détecter si du contenu est controversé ou non, ce qui nécessite une typologie de controverses
12. Un modèle pourrait être créé pour calculer la similarité entre un contenu généré par une IA et un contenu disponible en ligne et éventuellement sous droit
13. Un algorithme simple permettrait de répartir le revenu entre les différents auteurs, en fonction des données qui ont permis de formuler des requêtes aux utilisateurs.
14. Le modèle de langage lui-même pourrait être configuré pour reconnaître et signaler les informations personnelles dans les données d'entrée.
15. L'IA générative devrait signaler la source et le contenu sur lequel il se base pour répondre à une question.
16. L'IA générative devrait intégrer des filigranes dans les données générées et diffuser un programme permettant de vérifier si une œuvre contient le filigrane correspondant
17. Le modèle de partage devrait être implémenté dans le LLM lors de l'exécution des requêtes afin de pouvoir calculer des revenus sur une période donnée.
18. L'exploitant devrait définir la durée de la rémunération et sa périodicité
19. L'IA générative pourrait être entraînée pour reconnaître et signaler des mentions de copyright manquantes
20. L'IA générative pourrait être entraînée pour reconnaître et signaler les données provenant de Gmail et Google Search
21. L'interface utilisateur devrait inclure un analyseur de prompt permettant d'enregistrer la mention d'un nom ou d'un titre avant de le transmettre en entrée du LLM.
22. Les utilisateurs du modèle pourraient également contribuer à identifier des datasets controversés en signalant des problèmes ou des sources douteuses.
23. Les utilisateurs du modèle pourraient contribuer à identifier les suppressions / le manque des mentions de copyright en les signalant
24. Les hébergeurs de contenu devraient mettre à jour les conditions d'utilisation de leurs services
25. Obtenir le consentement des auteurs des données sur lesquelles le modèle est entraîné : mise à jour des paramètres de confidentialité
26. Une organisation collective pourrait être formée pour surveiller et évaluer l'origine des datasets utilisés dans les modèles de langage et créer une base de données de datasets d'origine frauduleuse.
27. L'interface utilisateur devrait avertir l'utilisateur de la potentielle inexactitude des informations fournies par le LLM
28. Les attaques par inférence d'appartenance (Membership Inference Attacks) permettent de prouver qu'un élément appartenait au corpus d'entraînement d'une IA générative. Ils constituent cependant un champ de recherche en pleine expansion et leur mise en œuvre reste complexe.

Les solutions techniques pour permettre la vérification des données utilisées pour l'entraînement des IAG sont à l'heure actuelle peu satisfaisantes pour permettre un déploiement efficace à grande échelle. La détermination des œuvres utilisées pour l'entraînement sur simple examen des paramètres ou des productions de l'IAG n'est pas suffisamment fiable pour acquérir la certitude qu'une œuvre donnée a été utilisée ou non pour son entraînement. Des techniques émergentes dans la littérature scientifique proposent la mise de place de marquages (ou « filigranes ») dans les

jeux de données dans le but de pouvoir tester leur présence ou non après entraînement des IAG. Ces techniques, bien qu'offrant quelques résultats encourageants pour les images notamment, présentent cependant des limitations qui seront selon nous difficiles à contourner : en particulier, nous ne sommes pas en mesure de déterminer (i) si ces techniques testées dans des situations isolées conservent leur efficacité à grande échelle, (ii) si ces techniques sont suffisamment efficaces pour protéger les œuvres de petite taille au regard de la grande quantité de données traitées

par les IAG, (iii) si ces techniques seront toujours valables lorsqu'une nouvelle architecture d'IAG apparaîtra. **Par ailleurs, de telles techniques ne sont pas rétroactives, et ne permettront donc pas de traiter la question des IAG déjà entraînés.**

Les difficultés que pose la mise en place de solutions techniques efficaces pour la protection des ayants-droits n'empêchent cependant pas les auteurs et les administrateurs des jeux de données de signaler leur refus que les œuvres et leurs données soient utilisées pour l'entraînement des IAG à l'avenir. L'interdiction de la récupération à des fins d'entraînement de certaines données en ligne est techniquement possible, et tout propriétaire de données qui ne souhaite pas que celles-ci soient utilisées devrait pouvoir activer un tel mécanisme, par exemple en manifestant ce refus via le fichier robots.txt pour les sites Web, comme prévu par OpenAI pour son programme d'indexation GPTBot<sup>(1)</sup>, ou en mettant à jour les conditions d'utilisation des plateformes à l'instar du New York Times ; le législateur pourra alors décider de la nature de la réponse à apporter en cas de non-respect de ces termes. Cependant, ce mécanisme n'est a priori valable que pour bloquer de grands acteurs. Il suppose de connaître les robots susceptibles de collecter des données<sup>(2)</sup>, et repose sur la bonne foi de ces derniers : du fait des difficultés techniques précédemment évoquées, le respect de cette interdiction dans les faits est invérifiable.

Par conséquent, nous estimons que la réponse la plus appropriée — et de surcroît la plus simple — serait de demander aux concepteurs d'IAG de faire preuve de transparence, en publiant notamment le détail des données utilisées pour l'entraînement de leurs produits. Cette transparence apporterait des solutions à la plupart des questions posées, puisque les œuvres utilisées seraient ouvertement déclarées ; une telle solution permettrait également d'adresser les différentes typologies d'IA et ne nécessiterait aucune adaptation pour suivre l'évolution des IA au fil du temps. La mise à disposition de ces listes de données faciliterait le calcul d'une éventuelle rémunération lors de demandes d'ayants-droits, qui serait alors faite au prorata des œuvres concernées dans le volume global. Le calcul pourrait ainsi être fait non pas sur le nombre mais sur la quantité de données numériques apportée par l'œuvre en question dans le jeu de données.

Il convient cependant de noter que cela impliquerait donc d'imposer le respect de règles internationales pour tous les concepteurs d'IAG, que celles-ci soient payantes ou *open-source*, et de faire en sorte que ces règles soient également applicables à tout autre outil à venir, ce qui paraît délicat. Le législateur français pourrait cependant imaginer de tels dispositifs techniques pour protéger les données sur le territoire français, par exemple via un « tag » générique dans un fichier de type **robots.txt**) que les concepteurs d'IAG s'engageraient à respecter, permettant là aussi au législateur de décider des actions à mener en cas de contournement de cette protection. Du fait des difficultés techniques posées par l'identification des œuvres utilisées après entraînement, cette approche s'appuie cependant sur un certain degré de confiance accordé aux concepteurs quant à leur engagement à respecter de telles règles, car s'assurer d'un contournement illégal pourrait s'avérer difficile face à un concepteur malhonnête.

En conclusion, il n'existe pas à ce jour de solution technique globale permettant une protection totale des ayants-droits. La piste de la transparence des concepteurs d'IAG nous apparaît comme la solution la plus simple à mettre en place pour le législateur, puisque celle-ci permet de passer outre les difficultés techniques évoquées. Il convient cependant de préciser deux limites à cette approche. Premièrement, nos préconisations ne couvrent pas le cas d'un utilisateur qui téléchargerait un LLM *open-source* non entraîné, puis qui l'entraînerait localement sur son ordinateur : cette démarche échappe *de facto* à toute forme de contrôle. De manière générale, nos préconisations ne concernent donc que les IAG conçues et entraînés par les grandes plateformes. Deuxièmement, les innovations technologiques actuelles offrent une large place à la génération de données synthétiques pour remplacer les données protégées par le droit d'auteur dans l'entraînement des IAG. Il faudra donc s'attendre, dans les prochaines années, à l'avènement d'une nouvelle génération d'IAG entraînés sur ce type de données. Les données synthétiques en question s'appuyant principalement sur les productions des IAG de première génération<sup>(3)</sup>, qui ont été entraînés avec des données protégées par le droit d'auteur, il se posera alors la question de déterminer la part de contribution de ces sources originelles dans l'entraînement des générations futures, et plus largement si le droit d'auteur sera toujours applicable à ces systèmes.

1. Voir [GPTBot - OpenAI API](#).

2. Voir [Comment ChatGPT et OpenAI pourraient utiliser votre contenu | DataDome](#).

3. Cette direction fait l'objet de recherches actives, y compris au sein de Talan. Voir l'article [« A Novel Incremental Learning Strategy Based on Synthetic Data Generated from a Random Forest »](#).

## RECOMMANDATIONS SUR LES MÉTRIQUES APPLICABLES

En prenant une liste de données utilisées pour l'entraînement d'une IAG, fournie par son concepteur selon le principe de transparence évoqué dans la partie précédente, nous préconisons une mesure de la part occupée par une œuvre concernée par une demande d'ayant-droit selon la quantité de données que cette œuvre représente dans le jeu de données complet. La mesure exacte de cette part dépend de la nature de l'œuvre considérée : dans le cas d'une image ou d'une vidéo, il est probable que cette part corresponde à la proportion de la taille de l'œuvre (mesurée en termes de pixels ou d'octets) dans la taille globale du jeu de données. Dans le cas des œuvres textuelles et des LLM, l'entraînement est basé sur un découpage du texte en *tokens*, qui ne correspondent pas exactement aux mots ni à un nombre précis de caractères, et donc d'octets. Si les informations sur l'algorithme de tokenisation sont fournies par le concepteur

de l'IAG, la part la plus juste serait celle de la proportion de *tokens* générés à partir de l'œuvre par rapport au nombre total de *tokens* générés pour le jeu de données complet. En l'absence de ces informations, un calcul approché mesurant la proportion de mots plutôt que de *tokens* nous paraîtrait être la solution la plus satisfaisante.

Idéalement, ces proportions pourraient être fournies directement par les concepteurs d'IAG en même temps que les listes de données utilisées ; néanmoins, que ces informations soient fournies ou pas, un organisme de gestion collective pourrait être chargé de collecter les jeux de données déclarés et effectuer directement la mesure de ces proportions, et le cas échéant confirmer ou infirmer les déclarations des concepteurs.

# SAJE : « LE JUGE DOIT POUVOIR ACCÉDER À LA LISTE DES «PROMPTS» DES UTILISATEURS, QUE CEUX-CI REVENDIQUENT UNE PATERNITÉ OU QU’ILS SOIENT SUSPECTÉS DE VIOLER LE DROIT D’AUTEUR »



Jean-Michel Salomon  
Président

## Objet : Contribution à l'expertise sur l'Intelligence Artificielle Générative (IAG)

Cher Monsieur Vincent Lorphelin,

Je vous remercie pour cet entretien très intéressant.

Je tiens à vous présenter une synthèse de mon point de vue, suite à notre discussion.

Dans le cadre de mon activité en tant que concepteur de jeux, j'ai été personnellement confronté à l'obligation de prouver ma qualité de créateur d'une œuvre.

Pour éviter toute contestation, mon avocate a procédé à un audit de l'historique des enregistrements du disque dur de mon ordinateur de travail. Cette démarche a permis de tracer d'une manière très précise le processus créatif, attestant sans équivoque de ma paternité sur cette œuvre.

Si les idées sont naturellement en libre parcours, leur expression représente le fruit d'un travail révélateur de l'intention et de la personnalité de leur auteur.

L'analyse du contenu du disque dur a rendu possible une reconstitution précise et claire du processus créatif, pour l'œuvre contesté intitulée « *Tout le monde veut prendre sa place* ».

En tant que représentant de la Société des Auteurs de Jeux (SAJE), je suis d'avis que le cadre législatif régissant l'Intelligence Artificielle Générative (IAG) devrait être élaboré

de manière à faciliter l'expertise sur laquelle se base le juge pour déterminer une rémunération équitable pour les auteurs. Plus précisément, il devrait être possible pour le juge d'obtenir aisément de la part de l'exploitant de l'IAG la liste des «prompts» des utilisateurs, que ceux-ci revendiquent une paternité ou qu'ils soient suspectés de violer le droit d'auteur.

L'œuvre générée par l'IAG repose également sur un historique influencé par ses utilisateurs. Il est alors possible d'en apprécier l'originalité.

Je vous prie d'agréer, Monsieur Lorphelin, l'expression de mes salutations distinguées.

Cordialement,

**Jean-Michel Salomon**

*Président de la SAJE*

*Société civile de gestion collective de droits d'auteur*

# BPCE : « NOUS ALLONS VOIR APPARAÎTRE DE SAVANTES FORMULES MATHÉMATIQUES POUR CRÉDITER ET RÉMUNÉRER LES AUTEURS »



Olivier Laborde

*Leader Innovation et Transformation digitale*

L'appétit vorace des systèmes d'intelligence artificielle pour les données se heurte aux limites du droit d'auteur. Pour s'entraîner et progresser, ces technologies doivent ingérer d'énormes quantités de contenus protégés. Or la loi interdit normalement la copie à grande échelle d'œuvres soumises au copyright.

Pourtant, grâce à la doctrine de l'« utilisation équitable ou usage loyal », les modèles LLMs (pour Large Language Models) se voient accorder une certaine latitude. Leur consommation de masse de chansons, images ou textes dans un but d'apprentissage automatique est tolérée, là où un humain commettrait une violation manifeste. Cette lecture robotisée échapperait ainsi au périmètre des lois de protection du droit d'auteur.

Cette exemption tacite, apparue presque par accident, a permis l'émergence des intelligences artificielles. Sans elle, impossible de constituer les immenses bases de données nécessaires à leur entraînement.

L'équilibre autour de l'« utilisation équitable » reste néanmoins précaire. À mesure que les systèmes d'IA prolifèrent, l'ampleur du recopiage interroge. Certains créateurs s'inquiètent de cette exploitation massive sans compensation. Des clarifications juridiques deviennent nécessaires pour concilier innovation technologique et droits des auteurs. Car même si l'insatiable soif de contenus des algorithmes échappe aujourd'hui aux lois de protection légales des œuvres, la position des entreprises d'IA n'est certainement pas tenable moralement.

Les tribunaux et les législateurs sont conscients des enjeux liés à l'utilisation croissante de l'IA dans la création artistique et cherchent à adapter les lois pour répondre à ces enjeux. Tout comme ils sont en train de regarder de près l'utilisation des données protégées par le droit d'auteur par les modèles pour leur entraînement.

Pour la clarification de ces deux situations, trois aspects doivent être pris en compte : le consentement, le crédit et la rémunération des auteurs.

## LE CONSENTEMENT DES AUTEURS, UN PRÉALABLE FONDAMENTAL

L'utilisation des textes par les intelligences artificielles génératrices de contenu pose un problème éthique majeur de consentement des auteurs. Nous l'avons vu précédemment, les modèles d'IA comme ChatGPT sont entraînés sur d'énormes quantités de textes glanés librement sur Internet, sans que les créateurs de ces contenus aient donné leur permission. Cette exploitation des œuvres sans autorisation

préalable va à l'encontre du droit moral fondamental des auteurs sur leurs propres créations.

Une solution concrète pour résoudre ce problème serait de mettre en place un protocole standard autorisant les auteurs à donner explicitement leur consentement, œuvre par œuvre, pour une utilisation éthique par les systèmes

d'IA. On pourrait imaginer un mécanisme simple d' « opt-in », où les créateurs choisiraient activement de participer à l'entraînement des intelligences artificielles en signalant les textes qu'ils acceptent de mettre à disposition à cette fin. Cela redonnerait aux auteurs un contrôle essentiel sur l'utilisation qui est faite des fruits de leur travail. Les initiatives comme le robots.txt d'OpenAI vont dans ce sens mais

restent beaucoup trop limitées : il faut donner aux auteurs une granularité fine pour choisir précisément, texte par texte, les contenus qu'ils acceptent de partager pour l'entraînement de l'IA ou pour la génération d'œuvres nouvelles assistée par l'IA. Le consentement éclairé doit être la pierre angulaire de l'éthique des données dans le domaine des IA génératrices.

## LE CRÉDIT AUX AUTEURS, POUR UNE NÉCESSAIRE TRANSPARENCE

Même quand les textes, les chansons et les images sont librement accessibles en ligne, les modèles d'IA génératifs devraient systématiquement créditer les auteurs dont ils s'inspirent et se nourrissent. Actuellement, des systèmes comme ChatGPT produisent des contenus de façon opaque, sans référencer leurs sources ni même signaler clairement qu'il s'agit de productions artificielles.

Pourtant, des techniques de « *citation driven* », référençant explicitement les sources textuelles utilisées, sont tout à fait

possibles à implémenter. Par exemple, on pourrait entraîner les IA à systématiquement citer leurs sources d'inspiration directement dans les contenus générés. Cette transparence sur la provenance des idées et formulations permettrait de résoudre en partie le problème éthique posé par les boîtes noires que sont aujourd'hui ces technologies. Le crédit aux auteurs originaux doit devenir la norme dans la conception des IA génératrices, il en va de la confiance des utilisateurs dans ces systèmes.

## LA JUSTE RÉMUNÉRATION DES AUTEURS, UNE NÉCESSITÉ MORALE

Enfin, dans les cas où les œuvres issues de l'IA sont exploitées commercialement, les auteurs des contenus ayant servi à l'entraînement et/ou à la création d'une nouvelle œuvre devraient aussi être rémunérés de manière équitable. À l'heure actuelle, les entreprises à l'origine de ces technologies ne versent aucune compensation aux créateurs dont elles tirent profit via l'exploitation des écrits, des images et des chansons.

Une solution serait de mettre en place des systèmes de redevances, sur le modèle des droits d'auteur qui régissent les industries culturelles traditionnelles. Même quand les textes sont ouverts et librement accessibles, leur exploitation commerciale à très grande échelle par les intelligences artificielles appelle à un juste accord avec les auteurs. Des accords sectoriels négociés pourraient être trouvés, ouvrant la voie à une répartition solidaire des revenus générés par ces nouvelles technologies reposant sur le travail humain antérieur. Le partage de la valeur doit être au cœur de l'éthique des IA génératrices de contenu.

Nous observons déjà quelques entreprises qui ont pris les devants et commencent à mettre en place des modes de rémunération des auteurs.

Adobe montre une voie au travers d'Adobe Firefly, une gamme de modèles créatifs de génération de contenu par l'IA disponible en tant qu'application Web autonome à l'adresse [firefly.adobe.com](https://firefly.adobe.com) ainsi que par le biais de fonctionnalités optimisées par Firefly dans les applications d'Adobe.

Le premier modèle Firefly commercialisé a été formé sur des images d'Adobe Stock, du contenu sous licence libre et du contenu issu du domaine public. Bien que les conditions qu'ils avaient déjà conclues avec les contributeurs Adobe Stock leur permettaient de créer et de former des modèles de Firefly, ils ont mis en place un programme de rémunération bonus Firefly, pour tous les contributeurs d'Adobe Stock éligibles dont le contenu a été utilisé dans la formation du jeu de données d'entraînement du premier modèle Firefly commercialisé.

Shutterstock, de son côté, a créé un fonds de contribution qui rémunèrera les contributeurs si leur contenu a été utilisée dans le développement de modèles générateurs d'IA, en plus du paiement de redevances liées aux futures licences de contenu généré par l'IA via l'outil de génération de contenu IA de Shutterstock.

Ascendant Art, lancé en mai dernier, agit de manière proactive, l'entreprise forme son IA (qui génère des images d'avatar), sur les images du domaine public, et promet de verser des redevances aux artistes qui soumettent volontairement leurs œuvres d'art sur lesquelles ses modèles pourront être formés.

Le développement responsable de ces technologies émergentes nécessite de trouver un équilibre entre les intérêts des acteurs de l'IA et le respect des droits des auteurs en termes de consentement éclairé, de crédit et de juste rétribution du travail créatif. La priorité doit être donnée à une

éthique morale et humaniste, plaçant les créateurs et leurs œuvres au centre des préoccupations.

En conclusion, même si l'entraînement des modèles d'IA génératifs s'avère couvert par l'« usage loyal », cela ne suffira pas à apaiser le débat et le sentiment que ces modèles utilisent des données et des œuvres protégées pour construire des modèles commerciaux. Dans cette optique, la question à résoudre est : quels remèdes et quelles techniques peuvent être introduits pour permettre à l'IA générative de s'épanouir tout en accordant du crédit et/ou une juste compensation aux créateurs dont le travail rend ce domaine possible ?

Pour y répondre, l'IA générative est certainement aussi la meilleure des chances, puisqu'elle peut tout aussi facilement tracker les plagiat que s'assurer de citer ses sources, voire générer les calculs de rémunération des auteurs qui auront été validés. Cela est techniquement possible et facile.

Les exemples d'Adobe, de Shutterstock et Ascendant Art sont de premières approches qui vont dans le bon sens. Les modalités d'utilisation des œuvres, la demande de consentement, le crédit et la juste compensation vont animer ces prochains mois et nous allons voir apparaître de savantes formules mathématiques pour créditer et rémunérer les auteurs.

Sans cela, [des outils de défense comme Nightshade et Glaze se développeront pour permettre aux artistes d'ajouter des modifications invisibles aux pixels de leurs œuvres](#) avant de les télécharger en ligne, de sorte que si elles sont intégrées dans un ensemble de données d'entraînement de modèles AI, le modèle résultant sera altéré et corrompu, causant de lourds dommages aux sociétés d'IA. Un véritable poison pour les LLMs.

# OKTAL-SE : « L'IDÉE SERAIT DE DÉFINIR UNE MÉTRIQUE AFIN DE QUALIFIER, EN FONCTION DES DONNÉES SOURCE, LE CURSEUR DE PLAGIAT »



**Jean Latger**  
*Président Directeur Général*

## LE TOUT SYNTHÉTIQUE

L'objectif de ce paragraphe est de montrer que l'IA Générative sera de plus en plus utile et utilisée dans le futur, non pas pour imiter l'homme et son talent créatif, mais pour entraîner des Réseaux Neuronaux à réaliser un nombre incalculable de fonctions dédiées à des domaines de plus en plus étendus. Je pense que, considérant l'ensemble de ces domaines, l'imitation d'œuvres d'auteurs devrait rester minoritaire.

OKTAL-SE utilise l'IA Générative dans un domaine très particulier du Traitement d'Images. Les produits d'OKTAL-SE dans l'IA permettent de générer des images de synthèse et constituer des datasets d'apprentissage pour des Réseaux Neuronaux dédiés aux fonctions de Détection et de Navigation. Par exemple, dans le secteur Défense la détection de cible et/ou la navigation vers un objectif. Par exemple dans le secteur aéronautique, l'aide à l'atterrissage et au taxiage, pour éviter les obstacles et suivre le balisage des pistes et taxiways. Par exemple dans l'automobile, le véhicule autonome et les Advanced Driver Assistance Systems pour la conduite semi-automatique.

Les datasets d'apprentissage sont la clef de voute de la capacité d'un RN à assurer sa fonction et à optimiser le scoring (taux de réussite) ainsi que l'intégrité (confiance en conditions critiques). L'enjeu est simple. Il s'agit de constituer des millions d'images (adaptées au contexte) avec « la solution » i.e. (avec le vocabulaire de l'IA) la labellisation, l'annotation, les tags, la segmentation, les métadonnées associés à l'image. Par exemple, pour l'automobile, les da-

tasets contiennent des images avec des piétons sur la route qui sont annotées au sens où on « dit » qu'il y a un piéton ou pas sur la route, afin de détecter des piétons traversants.

Avec des images générées synthétiquement, l'annotation des images est automatique. En effet, dans le cas des piétons par exemple, ce sont les outils d'OKTAL-SE qui positionnent virtuellement les piétons dans des maquettes virtuelles 3D ad hoc et qui les affichent dans l'image synthétique. Par conséquent, ces outils peuvent s'avoir où sont les piétons dans le champ de vue et comment ils sont potentiellement masqués par une entité de la scène 3D. Cet avantage est considérable. Le principal défaut est que ces images ne sont pas réelles. Le risque est qu'elles ne soient pas suffisamment réalistes. Ce sont des images simulées. Le danger est d'apprendre la simulation aux RNs et non la réalité.

C'est tout le challenge du moment : comment augmenter le réalisme (entropie) des images synthétiques pour pouvoir rivaliser avec la réalité.

Par ailleurs, il est important de prendre en compte plusieurs contraintes liées à ce marché de l'IA pour la détection et la navigation.

D'abord la problématique des capteurs. Une fois opérationnel et entraîné, le RN perçoit la réalité à travers des capteurs. Les principaux capteurs sont des caméras et des radars. Les caméras n'ont pas le pouvoir de résolution de l'œil humain mais ont beaucoup d'autres avantages. Par exemple, elles

« voient » en infrarouge la nuit. Les radars ont aussi leurs avantages. Ils « voient » à travers le brouillard et ils « perçoivent » la « vitesse » (Doppler) des cibles. Il n'existe pas de capteur universel. Par contre il existe de bonnes associations (fusion) de capteurs (multi-capteurs). Ceci laisse à penser que l'avenir est davantage le multi- capteurs qu'un unique capteur omniscient.

Ensuite la problématique des images réelles. Dans le cas d'images réelles, le travail d'annotation de million d'images reste très coûteux car fondamentalement manuel. De plus, à l'aune de la problématique capteur précédente, cette annotation devient techniquement très difficile. Il est en effet rare de trouver des images réelles de plusieurs capteurs colocalisés (monostatique), avec une unité de lieu et une unité de temps i.e. comme dans la réalité. Par exemple, les caméras et les radars imageurs ne « regardent » pas dans la même direction (squint en RADAR SAR). Si on prend l'exemple d'un satellite d'observation de la terre civil ou militaire, on peut difficilement garantir l'unité de lieu et de l'unité de temps i.e. « regarder » au même endroit et au même moment. C'est par contre le cas dans la réalité pour le RN embarqué. Le risque avec des images réelles « désynchronisées » est de créer des biais d'apprentissage (le cauchemar de l'IA) sachant que dans la réalité, l'environnement du RN embarqué est unique (par exemple les nuages n'ont pas bougé). Nous pensons chez OKTAL-SE que les images réelles ne sont pas la solution d'avenir, et donc que l'avenir est le synthétique et le génératif.

Par ailleurs, l'intégrité des RN est d'autant plus importante que les systèmes sont plus critiques. Par exemple, dans le domaine Défense, on attaque la nuit voire par temps de brouillard. C'est là précisément là que l'IA est la plus utile. Quand on fait l'acquisition d'images réelles (expérimentation) on recherche au contraire des conditions atmosphériques favorables.

Pour OKTAL-SE, dans son périmètre IA orienté détection et de navigation, l'avenir des RN pour ces applications, c'est le « quasi tout synthétique ». Le véritable challenge reste alors d'être capable de créer un véritable clone de la réalité, un « Reality twin » pour générer les datasets. Dans ces conditions, nous pensons que l'IA deviendra intrinsèquement générative dans le futur.

Ce qui milite encore plus pour le génératif pour ce type d'application, c'est le besoin de quantité d'images. Au-delà de cela, c'est de pouvoir traiter des cas aux limites qu'il serait très difficile de reproduire dans la réalité, typiquement dans le cas de conditions atmosphériques extrêmes ou plus simplement des points de vue particuliers (si on prend l'exemple de l'IA appliquée aux caméras intelligentes de

surveillance des foules, on trouve pléthore d'images réelles vues du sol et très peu vue « d'en haut »).

### **Autres données synthétiques**

Dans le monde du traitement d'image, les GANs (Generative Adversarial Networks) ont fait la preuve incontestée de leur efficacité. Le GAN combine deux RN, le « générateur » qui invente des images volontairement biaisées et apprend le « discriminateur » à se méfier, ce qui améliore ainsi son taux de réussite. Le « générateur » du GAN produit des données synthétiques qui n'ont aucune vocation bien évidemment à remplacer la main de l'homme. C'est une pure optimisation informatique.

### **Les données synthétiques et plagiat**

Le Copyright Office américain a défini une doctrine qui dédouane les contenus synthétiques de copyright et par suite dédouane aussi l'assemblage de données synthétiques de copyright.

### **La problématique du plagiat d'œuvres d'auteurs**

Le problème de légitimité du génératif pour l'IA est apparu dès les années 80 avec le travail de « David Cope » professeur de l'université de Californie qui créait de la musique de Chopin ou de Bach par apprentissage. Ce n'est donc pas nouveau.

De même, est-ce que les peintres de l'atelier de Leonard de Vinci, les Leornardeschi, étaient des plagiaires ? A priori non, ils peignaient à la manière de Leonard.

Les antiquaires, quand ils ne connaissent pas l'origine disent « à la manière de ».

La synthèse d'image c'est un peu comme de la « peinture ». Par exemple, pour créer des scènes virtuelles 3D représentant le paysage, on part de photographies aériennes, qui n'ont rien d'artistiques mais qui ont été largement retouchées par le « photographe » IGN (ortho-rectification, suppression des nuages, harmonisation des couleurs, etc.). Les modeleurs 3D qui créent une sorte de « CAO du paysage » amènent aussi leur touche. C'est l'infographiste le « peintre » et l'outil modeleur 3D la « palette » et les « pinceaux ».

La scène 3D est constituée de triangles « tapissés » par des textures issues de photos (StreetView, etc.). Est-ce que les textures, les palettes de couleur, les matériaux, doivent rétribuer « l'artiste d'origine » ?

Est-ce que les papiers collés de Georges Braque ou les collages d'Andy Wharhol sont l'œuvre de plagiaires ?

Pour mon Directeur Scientifique IA, et je partage cette idée, l'IA Générative se contente de faire des « collages intelligents », mais il faut bien sûr que ce soit un collage et pas une simple photocopie.

Nous pensons qu'à chaque génération, et, quel que soit le domaine, il y a, à chaque fois, un élément majeur différenciant, typiquement, le sujet, la mélodie, la taille, la forme etc.

## COMMENT QUANTIFIER LA PART DE PLAGIAT ?

Pour les cas litigieux où l'IAG est suspecte de plagiat avéré d'artistes, où par exemple on n'aurait aucun assemblage mais un simple copié/collé d'une œuvre, où par exemple il n'y aurait aucun élément différenciant de sujet ni de situation, il nous paraît nécessaire de réagir en mesurant et en quantifiant le niveau de plagiat.

L'idée que l'AI pourrait être utilisée, a posteriori, pour estimer la part de plagiat d'une IAG n'est pas la bonne voie selon moi. Je pense qu'il vaut mieux traiter le problème a priori et à la source.

Si les données à la source n'ont pas de copyright, leur assemblage, même simple, n'est pas du plagiat. Si les données à la

## CONCLUSION

L'approche d'OKTAL-SE en IA ne concerne pas les métiers artistiques. Dans ce domaine de l'image de synthèse, les contenus synthétiques ne sont pas sujets à copyright.

L'approche d'OKTAL-SE pourrait s'étendre aux métiers artistiques. En effet, les ortho-images satellitaires ne sont pas des œuvres d'Art, du moins jusqu'à aujourd'hui. Néanmoins OKTAL-SE utilise ces ortho-images satellitaires pour

Donc notre position serait que s'il existe un élément majeur de différence alors il n'y a pas de copyright.

Le papier « Research on Copyright of AI Copycat Works » de l'université Chinoise Lanzhou donne un éclairage intéressant sur les problèmes de copyright avec l'IAG : [https://www.shs-conferences.org/articles/shsconf/pdf/2023/06/shsconf\\_essc2023\\_04005.pdf](https://www.shs-conferences.org/articles/shsconf/pdf/2023/06/shsconf_essc2023_04005.pdf)

source ont des copyrights, mais qu'elles sont suffisamment parcellaires d'une part et que leur assemblage est suffisamment complexe d'autre part, alors ce n'est pas du plagiat.

L'idée serait de définir une métrique c'est-à-dire définir un espace et une mesure de distance dans cet espace, afin de qualifier, en fonction des données source, le curseur de plagiat (une valeur flottante et pas binaire). Pour ce faire, je crois qu'il faut d'abord clairement définir les espaces (peinture, musique, littérature, cinéma, jeu vidéo, images de synthèse, etc.) et puis définir, pour chacun de ces espaces, une métrique spécifique adaptée à ces différents espaces.

entraîner des RNs à « peindre » la réalité, mais pas à des fins artistiques.

Comme cette approche n'est définitivement pas du plagiat, si on étend cette approche au domaine artistique, cela laisse à penser que l'IAG n'est que très minoritairement du plagiat, même dans le domaine artistique.

# JOLIBRAIN : « DES POINTS DE MESURE EXISTENT DÉJÀ, LES CONCEPTS D'ŒUVRE ESSENTIELLE ET DE CORPUS NATIONAL PEUVENT ÊTRE CONSIDÉRÉS »



**Emmanuel Benazera**  
Président Directeur Général

De longs travaux d'un sous champ de l'Intelligence Artificielle, l'IA générative (IAG), a émergé un ensemble d'applications pour la production automatique de texte, images, sons, code informatique et autres objets numériques. Les productions de ces applications deviennent en pratique indiscernables de réalisations humaines. Et pour cause, elles en dépendent fortement.

L'IAG opère une prise d'empreinte de la distribution des données à l'échelle mondiale, qu'elles soient textuelles, visuelles ou autre. Lorsque l'empreinte mixe les types de source, l'on parle d'IAG multimodale. L'IAG n'a pas d'intention, et exécute un récital numérique, possiblement non déterministe, permettant de disposer pour chaque entrée de plusieurs sorties, mais en nombre théoriquement fini. Son pilotage est très flexible pour orientation et ciblage du contenu de la sortie générée. L'intention est donc externe, on parle techniquement de conditionnement du modèle d'IAG.

Il apparaît donc légitime d'étudier dans ses formes applicables la mesure de cette dépendance de l'automatisation des réalisations humaines aux réalisations humaines elles-mêmes, ainsi que ses conséquences économiques sur les acteurs en présence.

L'émergence de l'IAG reposant sur des éléments éminemment techniques et théoriques, en les considérant dans leur détail, il est possible de faire émerger des solutions simples et réalistes de mesure et d'équilibrage des revenus des acteurs.

Mais il faut d'abord bien considérer et voir la structure du champs IAG, composé de trois acteurs principaux. D'abord les producteurs des modèles d'IAG, en grande majorité des sociétés privées. Si leur taille est variable en effectif, l'IAG présente la particularité de reposer sur un très petit nombre de personnes et un gros volume de calcul. Il suit du coût de ce dernier<sup>(1)</sup> que leur capital financier et leur dépendance économique aux GAFAs est caractérisée<sup>(2)</sup>. OpenAI en est l'acteur emblématique, particulièrement pour sa dominance via les outils GPT-4 (modèle de langage pour génération de texte) et Dalle-3 (génération d'images), mais aussi par son nom en trompe l'œil, car ce qui avait commencé comme une aventure teintée d'ouverture, l'a transformé le succès venu en l'acteur le plus fermé du marché<sup>(3)</sup>. Deux dimensions principales structurent les positions des acteurs de production d'IAG. D'une part l'accès aux ressources de calcul, dont le volume détermine la dominance sur les autres acteurs. Ainsi, OpenAI, Microsoft, Meta et Nvidia dominent les autres, repoussant les startups dans une position de dépendance

1. Le coût par modèle est estimé entre 10M\$ et 40M\$, et la location de 4000 GPUs en « cloud » est estimée à 40M\$ par an. De ce fait une levée de fonds de 100M\$ correspond à deux ans de calcul environ avec 4000 GPUs.

2. Les acteurs privés de taille moyenne, comme Anthropic.AI, Cohere.AI et OpenAI voit leur activité reposer en pratique sur des investissements récurrents de Google, Microsoft, Nvidia ou Salesforce. Seul Midjourney, le petit poucet de l'IA annonce être rentable et sans investissement extérieur.

3. Les revenus d'OpenAI sont estimés à 100M\$/mois, ce qui correspond au financement de 3 à 10 modèles par mois.

dangereuse. Nvidia est un acteur particulier pour le calcul, puisque ses propres investissements au capital des acteurs producteurs d'IAG lui reviennent presque immédiatement en revenu d'achat des processeurs GPUs qu'il produit<sup>(4)</sup>. La seconde dimension est celle de l'ouverture ou non des modèles, on parle, parfois à tort en termes de définition stricte, de modèles open source, par opposition aux modèles seulement disponibles derrière un accès contrôlé. Meta se fait le champion des modèles open source, avec quelques acteurs émergents, et notamment français, comme Mistral. AI. La disponibilité de modèles open source est cruciale nous le verrons pour la mesure et la rémunération des ayants-droits.

Ensuite viennent les utilisateurs des produits IAG, allant des citoyens, indépendants et artistes aux acteurs de l'ingénierie, du marketing et de la publicité pour n'en citer que quelques-uns. Les axes structurants sont d'une part celui de la dépendance au champs économique et d'autre part celui de la création autonome, presque avant-gardiste. Ces sujets en France font l'objet d'études approfondies<sup>(5)</sup>. Dans ce cadre, l'IAG abaisse les coûts de création de tous les acteurs, aussi bien sur le marché des biens symboliques, que sur celui du code informatique par exemple. Cet état de fait favorise néanmoins en pratique les acteurs déjà puissants, et notamment de diffusion de contenu en masse (marketing, publicité, ...) pour lesquels la baisse des coûts est drastique<sup>(6)</sup>. Pour les autres il s'agit d'un accélérateur du travail, surtout dans les professions du numérique.

L'effet de renforcement des inégalités est double, puisque cette baisse de coût pour les utilisateurs dominants du champ économique se fait au détriment des acteurs de la production de contenu, qui sont le troisième acteur de

l'IAG. En effet, en pratique, les modèles construits par les entreprises informatiques du privé reposent sur l'ingestion en masse, et bien souvent indiscriminée, de tous les contenus disponibles à l'échelle du Web, c'est à dire mondiale, et au-delà via des silos privés. Les acteurs de la production de contenu se retrouvent donc à la fois utilisateurs (c'est à dire consommateurs de l'IAG), et en quelque sorte victimes puisque leurs créations servent à nourrir des modèles remis à jour en continu, c'est à dire englobant la capacité à produire ce qu'eux-mêmes viennent de produire.

Paradoxalement, à ce stade de la technique, les producteurs de modèles IAG apparaissent déficitaires<sup>(7)</sup>. De ce fait, étant donnée la structure de dominance de ces producteurs privés, toute tentative d'élaborer par la mesure et la redistribution un rééquilibrage durable des positions et des forces doit se faire de façon réaliste, en prenant en compte d'une part les réalités techniques de l'IAG, et d'autre part les velléités de rétention attendues de la part des principaux acteurs, en cherchant à les contourner.

Pour ce faire, nous suggérons trois points de mesure réalistes, permettant l'établissement par la suite de règles de rétribution, dans des périmètres d'applicabilité qui sont ensuite de fait à considérer par le législateur. Ces trois points d'entrée sont dans l'ordre :

- Les corpus servant à la construction des modèles d'IAG ;
  - Les modèles eux-mêmes avec leur spécificité de structure et d'opacité ;
  - Leur usage, c'est à dire dans le conditionnement et l'intention déjà citées permettant l'orientation et le ciblage des résultats.
- 
- Des conversations, échanges, et éléments courts en milliards et au-delà issus des réseaux sociaux ;
  - Des corpus privés, achetés à des ayants-droits, que ce soit dans les domaines juridiques, techniques...

## CORPUS

Les corpus d'apprentissage sont composés de milliards de documents, textuels, visuels ou autre. On peut considérer que ceux-ci dans leur grande majorité proviennent des sources suivantes :

- Des œuvres publiquement diffusées, dont seulement certaines ne font pas ou plus l'objet de restrictions en termes de licences, de copie ou de rediffusion ;

4. A moyen terme, il est attendu une baisse continue des coûts de calcul, donc de réalisation et d'exploitation des IAGs, avec possiblement une hausse drastique du coût de l'électricité en tension sur plusieurs fronts.

5. Pierre Bourdieu, *Les règles de l'art*, 1992. Et plus particulièrement le chapitre 3 traitant du marché des biens symboliques.

6. [https://www.liberation.fr/economie/social/chez-onclusive-un-des-premiers-plans-massifs-de-licenciement-dus-a-lia-en-france-20230914\\_WMEUA-46RHRCRMV15HPZJXSD4A/](https://www.liberation.fr/economie/social/chez-onclusive-un-des-premiers-plans-massifs-de-licenciement-dus-a-lia-en-france-20230914_WMEUA-46RHRCRMV15HPZJXSD4A/)

7. L'exploitation des IAGs reste majoritairement déficitaire, OpenAI et Microsoft, ce dernier via la plateforme GitHub opérant l'IAG pour le code « copilot », rapportent des pertes importantes, cf. <https://www.wsj.com/tech/ai/ais-costly-buildup-could-make-early-products-a-hard-sell-bdd29b9f>

La dynamique des sources des corpus diffère. Ainsi on peut considérer que la masse de données quotidienne issue des réseaux sociaux domine très largement les autres sources, pouvant être considérées comme statiques en comparaison.

D'autre part, la rémunération de ces corpus diffère aussi. En général les conditions d'utilisation des réseaux sociaux régissent un système d'échange de service gratuit contre utilisation des données, mais pas toujours. Ce qui est très certainement observé depuis l'avènement des produits IAG est la fermeture par les principaux réseaux sociaux de l'accès gratuit à leur contenu par des tiers<sup>(8)</sup>, y voyant un moyen de rémunération d'activités historiquement déficitaires.

Ces corpus font l'objet d'un énorme travail de nettoyage, de déduplication et de purification. Si l'IAG repose bien sur des principes techniques mathématisés, ceux-ci échouent en pratique à réaliser de facto des produits puissants et contrôlables si les données ne sont pas très finement sélectionnées et pré-travaillées.

Techniquement, la phase de construction du modèle IAG par apprentissage machine se fait en deux à trois étapes, dont les plus structurantes sont :

- Une étape d'apprentissage sur données en masse ;
- Une étape de façonnage du contrôle du modèle, pour favoriser son pilotage par les utilisateurs, et moduler certaines sorties non désirées.

Différentes variantes sont utilisées pour la génération de texte ou d'images typiquement. Mais si le fond technique diffère, modèles autorégressifs pour la génération de texte, modèles dits de débruitage par diffusion pour la génération d'images, les grandes étapes sont similaires, et ne nous semblent pas impliquer de différences de conséquence en ce qui concerne les mesures possibles d'influence des contenus par exemple.

## Mesures

Au sens de la construction du modèle d'IAG, la composition du corpus apprêté représente de facto la composition apprise par le modèle. Cet effet est renforcé par le fait qu'en pratique la construction s'effectue par un passage unique sur le corpus<sup>(9)</sup>. A notre connaissance il n'y a pas de renforcement ou passage multiple effectué sur des sous-parties du corpus au détriment des autres, mais toutes les recettes ne sont pas rendues publiques. De plus, les techniques existantes, cad. grossièrement l'apprentissage

automatique profond via réseaux de neurones, se révèle loin en pratique des performances humaines en termes de ratio de performance par contenu ingéré<sup>(10)</sup>. Une conséquence est l'ingestion de corpus exhaustifs, notamment à l'échelle de l'Internet, car la donnée même à cette échelle constitue de facto une variable limitante. Ce qui sous-entend que toutes les données disponibles peuvent être considérées comme présente dans les corpus de construction des IAGs.

Il suit de ces premières remarques que l'influence des contenus du corpus se fait peu ou prou au ratio de leur présence dans le corpus global, et qu'il n'est pas nécessaire de les marquer, puisque par hypothèse réaliste, elles y sont toutes.

De ce fait les corpus, par leur structure de source et leurs caractéristiques établies ci-dessus, permet une mesure approximative simple :

- Les œuvres publiquement disponibles sont connues et listées (par ex. « library of congress », ISBN, ...) ;
- Les données disponibles sur les réseaux sociaux sont publiques, même si pas toujours accessibles en masse exhaustives pour les raisons suscitées ;
- Les données privées disponibles à l'achat ont un prix, connu.

Il est donc possible et simple de réaliser des approximations de présence des contenus et des ayants droits, extrapolables là encore relativement facilement. A noter la présence en masse de données issues des réseaux sociaux, qui peut légitimer sur certains pans, la notion de « tous producteurs ».

## Recommandations

Sur ces bases, nos recommandations relatives à la mesure et aux actions réalistes de rééquilibrage des droits suivent :

1. Les sources des corpus permettent une mesure approximative extrapolable en considérant que toutes les œuvres y sont, à hauteur de leur présence dans les corpus accessibles ;
2. La constitution d'un corpus ou de plusieurs corpus nationaux d'œuvres et contenus. Ceci est un point crucial, car l'élaboration des corpus par les acteurs privés est un travail caché et de très haute importance. En produisant ses propres corpus, la France pousse son catalogue caractérisé et contrôlé vers les acteurs de production d'IAG. Cela simplifie leur travail, abaisse leurs coûts, nivelle la concurrence sur les corpus, assure que les œuvres

8. Ainsi Reddit et Twitter ont fermé leurs accès, et l'ouvre moyennant rémunération par les producteurs d'IAG. Reddit et Twitter sont déficitaires.

9. On parle d'époques de l'apprentissage, une époque correspondant à une passe complète sur le corpus.

10. <https://www.lesswrong.com/posts/t6MntQzsdY66ayDf4/how-does-the-human-brain-compare-to-deep-learning-on-sample> pour les LLMs.

françaises soient bien représentées<sup>(11)(12)</sup>. Via ces corpus nationaux les ayants-droits voient leurs droits traités en amont de la production de l'IAG et stabilisés (le modèle IAG change, le corpus aussi mais indépendamment, en amont).

3. La constitution de corpus sous licence Creative Commons (CC), permettant de facto la construction de modèles libres de droits. Ces corpus permettent, nous le verrons, l'établissement de certaines propriétés des modèles et de leur conséquence sur les ayants-droits.

En résumé, la constitution de mesures sur corpus nationaux est cruciale, et permet l'établissement de règles de rétribu-

## MODÈLES

Le modèle IAG, résultat de la construction par apprentissage machine, peut être intuitivement compris comme une démocratisation des contenus via leur compression pour n'en conserver que leur « substantifique moelle ». C'est à dire leur valeur profonde, permettant à la machine de les recomposer, ou plus exactement, de les réciter. On parle de « généralisation » d'un modèle lorsque la compression qu'il opère lui permet de remplir par ses sorties l'espace des contenus situé entre les contenus fournis dans le corpus d'apprentissage, on peut aussi parler d'interpolation avancée. Dans ce sens, tout apprentissage est compression, et l'intelligence serait une forme de compression de l'exhaustivité du réel. Le modèle d'IAG opère donc une compression culturelle proche de celle réalisée par les humains, mais à échelle plus large, et créativité moindre.

Lorsque le modèle est dit « open source », en pratique cela signifie que sa représentation sous forme de fichier numérique, contenant les « poids » de sa structure sous forme de réseau de neurones, sont disponibles publiquement. Si la licence est suffisamment permissive, le modèle peut être reciblé sur une tâche déterminée. Par exemple par une phase d'apprentissage court sur un corpus spécialisé, réduisant ainsi le domaine d'application du modèle initial, à l'avantage de sa précision sur la nouvelle cible.

Il faut bien noter que qu'il soit open source ou pas, le modèle, par les repliements complexes qu'il effectue sur les données, et qui sont les opérations élémentaires non-linéaires de la compression drastique qu'il opère, ne peut que malhabilement être expliqué et décortiqué en pratique. En d'autres termes, l'extraction de règles de dépendances

et d'équilibrage réalistes indépendamment des acteurs dominants de la production d'IAG.

Les acteurs dominants de l'utilisation en masse, c'est à dire les producteurs les plus dépendants du champ économique, principalement pour la production d'objets numériques éphémères à cycle de production court (par ex. marketing, publicité, ...), bénéficiant des baisses de coûts substantielles attribuées à l'utilisation de l'IAG, pourraient être mis à contribution afin de soutenir en partie la création des corpus nationaux.

simples entre ses sorties et ses entrées, dans ce sens, est illusoire et à faible précision en pratique, même si certaines garanties théoriques et algorithmiques peuvent exister sous certaines hypothèses. En pratique ces dernières ne sont pas réalisées, et nous suggérons que ces approches ne sont pas suffisamment informatives, trop coûteuses en calcul et en temps, et relativement peu fiables. Nous ne comptons pas sur leur amélioration drastique dans un futur proche, par structure même de la compression extrêmement complexe opérée par les modèles.

Il en suit que la copie conforme est en pratique impossible dans l'absolu, voire involontaire (bien que le cas demeure improbable au sens statistique). Néanmoins le modèle conserve la possibilité de rendre copie approchée, voire même assez précisément, d'une œuvre, via certaines manipulations, ou conditionnements intentionnels. Un audit pour copie conforme au pixel près pourrait prendre cependant un temps quasi infini.

La question devient donc de savoir si et quelle influence une œuvre donnée peut avoir pour le modèle, cad. en pratique sur ses sorties. Cette question peut être appréhendée scientifiquement via l'hypothèse dite « de variété » (en anglais « Manifold Hypothesis »). Celle-ci stipule que l'ensemble des données (par exemples des visages, ou des paysages, ainsi que leur combinaisons) reposent sur un support, c'est à dire grossièrement un sous-ensemble du paysage de tous les contenus possibles, qui est bien inférieur en taille et en dimensions à l'ensemble des possibles lui-même. Ce support est par ailleurs relativement compact, c'est à dire qu'il prend peu d'espace.

11. C'est considération est sérieuse, des masses de contenus en Chinois viennent d'être ouverts aux producteurs d'IAG exactement dans ce but, afin d'éviter une sous-représentation face aux contenus en Anglais, voir <https://annas-blog.org/duxiu-exclusive.html> publié le 4/11/2023.

12. De premiers efforts limités existent en France, <https://www.loria.fr/fr/le-projet-llm4all-finance-par-lanr-un-consortium-gagnant-apportant-les-der-nieres-avancees-en-matiere-dia-generative-au-service-de-la-sante>

Autrement dit, que l'ensemble des contenus et œuvres réellement réalisables, ou encore, présentant une structure appréhendable pouvant évoquer autre chose que du bruit, reposent les unes à côté des autres dans un mouchoir de poche. Cette hypothèse est validée empiriquement de plusieurs façons, notamment via l'étude des positions des contenus au sein des espaces comprimés par les réseaux de neurones.

En pratique, il vient que d'une part tous les contenus sont dépendants les uns des autres car leur proximité dans l'espace implique des fondements et supports quasi-identiques, d'autre part qu'il existe à peu près forcément des contenus essentiels, dont la particularité est d'étendre plus que les autres la surface du support.

## Mesures

Sur ces bases, nous présentons un ensemble de mesures rationnelles et qui nous semblent scientifiquement valides :

- L'existence, et même l'accumulation de modèles « open source », permet d'envisager tout un ensemble de mesures de présences des œuvres si nécessaire, d'audits pour copie conforme s'il le faut, et d'analyses. Celles-ci pourraient dès lors être transposées de facto aux modèles fermés. L'un des avantages pourraient être de pousser les producteurs d'IAG à proposer de facto des versions moindres ou bridées de leurs modèles mais « open source » pour analyses approximées mais extrapolables à leurs modèles fermés ;
- L'existence d'Œuvres Essentielles, cad. jouant un rôle unique et privilégié vis à vis des autres œuvres, notamment en termes de dépendances de ces dernières vers l'Œuvre Essentielle. Cette existence est rationnelle et observable, bien que certainement difficile voire impossible à quantifier ;
- L'utilisation de corpus sous licence CC permettant d'envisager des modèles sans droits inclus. De tels modèles

## USAGES

Les IAGs permettent la production en masse et à vitesse intensifiée de contenus devenus en pratique indiscernables des autres contenus non automatiquement générés.

L'utilisation se fait ou bien de façon complètement aléatoire, la machine générant un contenu sans intention, ou bien de façon conditionnée, cad. via l'intention d'un utilisateur sous forme de texte descriptif, image d'entrée, etc. En pratique c'est cette seconde utilisation qui domine très largement, et donc celle que nous étudions.

A chaque conditionnement intentionnel va correspondre un ensemble fini de sorties de l'IAG. Ces sorties peuvent

peuvent permettre d'estimer si la machine récitante peut générer une copie quasi-conforme d'une œuvre, indiquant de ce fait que l'œuvre est comprise dans un corpus CC, et par conséquent probablement essentielle car contenue dans les autres.

## Recommandations

Les modèles, même « open source », sont en pratique opaques. Néanmoins il nous est possible de faire certaines recommandations réalistes, là encore indépendantes du bon vouloir des producteurs privés d'IAG :

1. La constitution de modèles nationaux sur base des corpus nationaux déjà recommandés. Ces modèles permettraient des mesures et audits tout en contrôlant le périmètre des ayants-droits. Ces modèles peuvent être établis par la recherche académique en France et à l'international, bénéficiant ainsi à la recherche fondamentale et appliquée, ainsi qu'au champs culturel tout entier puisque les modèles seraient améliorés de ce fait par l'état de l'art du champs IAG qui bouge très vite actuellement ;
2. La notion d'Œuvre Essentielle (OE), obtenue non pas par quantification hasardeuse ou même via des outils mathématiques, mais par labellisation par le champ autonome qu'est le marché des biens symboliques. Il est tout à fait prévisible qu'en pratique les OEs vont se trouver dans la partie du champ la moins dépendante du champ économique, c'est à dire celle de l'art pour lui-même, par opposition à celle de la diffusion de masse via production en cycle très court (marketing, publicité, etc. suscités). Cela permettrait de contrer un peu le double renforcement des inégalités dû à l'utilisation des IAGs, tout en singularisant des éléments marquants du corpus national.

être modifiées et/ou réinjectées en entrées, constituant une séquence intentionnelle interactive avec l'IAG. Cette séquence, et la nature de la récitation ou de l'interpolation des contenus existants par l'IAG, sans intention, sous-entend que toute nouvelle œuvre générée peut être comprise comme produite par l'intention, et donc sujette à droit. Au législateur d'en déterminer les contours exacts.

## Mesures

Dans ce sens, deux points de mesures émergent naturellement : le conditionnement (aussi appelé « prompt » en anglais lorsqu'il s'agit de texte), et la sortie générée. En

pratique il nous semble que ces mesures peuvent être effectuées de façon relativement simple et utile :

- Les conditionnements et « prompts » peuvent être analysés comme nécessaire. Leur obtention en provenance des producteurs privés serait un plus mais n'est pas nécessaire à notre avis. En effet, la diffusion publique des conditionnements est telle que des sites et entreprises spécialisés existent et diffusent cette connaissance. Celle-ci peut donc être observée en ligne sur le Web ou bien achetée à des tiers. Il semble réaliste que les générateurs de contenus largement éphémères, en masse, et au plus faible coût se reposent sur ces savoir-faire. L'audit de ces données disponibles apparaît ensuite tout à fait extrapolable au reste du marché des usages ;
- Les sorties des IAGs accompagnent pour la plupart toute sorte de publications. Dans ce sens, ces sorties sont elles aussi disponibles en masse sur le Web et les réseaux sociaux en particulier. Leur détection et audit n'offre donc pas de grande difficulté technique, et reste extrapolable statistiquement aux autres sorties ;
- Le corpus national déjà recommandé doit permettre de servir aussi de base de référence des œuvres déposées par les ayants droits. Des algorithmes de recherche par similarités, qui sont dans leur ensemble bien maîtrisés, peuvent être facilement utilisés pour singulariser et quantifier les possibles copies quasi-conformes ou dérivations à rechercher. Il n'y a pas de verrou technologique, et la notion de corpus national prend encore ici tout son sens pratique, s'il n'y avait que cela.

## Recommandations

Nos recommandations suivent les structures décrites ci-dessus en ce qui concerne l'usage des IAGs en pratique :

1. Gratuité de l'utilisation de modèles nationaux construits sur corpus nationaux, aux ayants-droits ;

2. Financement des corpus et modèles nationaux via l'usage ou la contribution des producteurs de contenus de masse. On pourrait y inclure le financement de datasets sous licence CC, opérés par des structures associatives dédiées ;
3. Rémunération des ayants-droits via mesure d'usage sur données publiques.

En résumé, il nous semble que le rééquilibrage présent et à venir des changements liés à la création de contenus accélérés par l'IAG peut tout à fait se faire en évitant l'écueil immédiat d'une dépendance aux producteurs privés d'IAG. Ceci car non seulement les solutions sont disponibles, mais aussi parce que ces producteurs font encore face à des coûts de calcul hors norme, sans rentabilité immédiate, ainsi qu'une compétition féroce, les rendant peu aptes à collaborer rapidement.

Or les notions de corpus et modèles nationaux d'une part, et d'œuvre essentielle d'autre part, permettent d'envisager un positionnement positif, singulier et utile pour un rééquilibrage nécessaire à venir.

Ces deux propositions non seulement mettent en valeur le patrimoine culturel national, ainsi que sa dynamique, mais aussi les positionnent comme parties intégrantes d'un changement technique qui n'est au final, qu'un nouveau pinceau dont les générations présentes et futures sauront très bien se servir, pour peu qu'on leur en donne l'occasion positive et sécurisée.

# GOLEM.AI : « DES MODÈLES ALTERNATIFS À L'IA GÉNÉRATIVE DOIVENT ÊTRE PRIVILÉGIÉS POUR MAINTENIR UNE TRAÇABILITÉ TOTALE »



**Killian Vermersch**  
*Co-Fondateur et Directeur Général*

Un algorithme génératif de textes ou d'images est basé sur un grand nombre de données entraînées pour la constitution de son modèle.

Les modèles actuels LLM sont constitués en moyenne de 70 milliards de paramètres pour les plus grands modèles et de 7 milliards de paramètres pour les plus "légers"

Ces paramètres sont des éléments qui sont créés lors d'un processus d'entraînement, qui peuvent représenter des biais, poids, etc. qui sont invisibles dans le raisonnement d'une IA basée sur des réseaux de neurones (effet boîte noire).

Il est donc techniquement impossible de déterminer la source précise d'un contenu généré.

Il est plus simple de consulter les sources utilisées lors de l'entraînement, que de chercher à corréliser les inspirations utilisées dans une réponse d'IAG.

Le monde de l'intelligence artificielle est très hétérogène, que ce soit dans les usages ou les techniques utilisées. Dans le cadre de notre problématique, la séparation par approche semble la plus pertinente :

- **L'approche connexionniste ou application statistique :**

Cette approche la plus connue utilise des modèles de données, issus d'entraînement afin de répondre statistiquement

à une question. C'est dans cette famille que se trouvent les LLM utilisés par l'IA Générative, le machine learning, le deep learning, etc. C'est dans cette approche que le problème de traçabilité et d'explicabilité des décisions est présent.

- **Approche symbolique ou l'application d'une science dure :**

Cette approche est l'application algorithmique des sciences. Ces intelligences artificielles sont souvent qualifiées de systèmes experts, systèmes ontologiques, programmations logiques, réseaux sémantiques, etc. Par nature, ces algorithmes sont explicites, démontrables et plus légers en consommation qu'une approche connexionniste. L'explicabilité de ces algorithmes permet de tracer la source de l'information qui a été utilisée pour une analyse.

En conclusion, dans le cadre des intelligences artificielles statistiques, la simple présence d'une œuvre dans les sources de données d'entraînement doit compter pour l'application du droit d'auteur.

Si on souhaite avoir un modèle plus équitable liée à l'utilisation des œuvres dans les réponses fournies, alors des modèles alternatifs à l'IA générative doivent être privilégiés pour maintenir une traçabilité totale.

# FRANCE META : « LA RÉMUNÉRATION ÉQUITABLE DANS LE DOMAINE DE L'IA GÉNÉRATIVE NE CONCERNE PAS SEULEMENT LES AUTEURS INDIVIDUELS, MAIS TOUCHE ÉGALEMENT À LA SANTÉ GLOBALE DE L'INDUSTRIE DE L'IA ET À L'ÉQUITÉ AU SEIN DE LA SOCIÉTÉ »



Emmanuel Moyrand  
Co-Fondateur

## IMPACT DE LA RÉMUNÉRATION ÉQUITABLE : DYNAMISER LE SECTEUR DE L'IA PAR LA JUSTICE ET L'INNOVATION

L'instauration d'une rémunération équitable pour les auteurs de données utilisées dans l'entraînement des modèles d'intelligence artificielle (IA) générative peut avoir des répercussions profondément positives sur l'ensemble du secteur. Voici une analyse des impacts bénéfiques d'une telle mesure :

- 1. Stimulation de l'Innovation :** En rémunérant équitablement les auteurs pour leurs contributions, nous encourageons une plus grande créativité et innovation. Les créateurs de contenu seront plus enclins à partager leurs œuvres pour l'entraînement des IA, sachant qu'ils seront justement récompensés. Cela conduit à un afflux de données plus diversifiées et innovantes, ce qui est essentiel pour le développement d'IA plus performantes et ingénieuses.
- 2. Amélioration de la Qualité des Données :** Une rémunération équitable peut améliorer la qualité des données disponibles pour l'entraînement des IA. Les auteurs ayant un intérêt financier seront plus motivés à produire et maintenir des données de haute qualité. De plus, la reconnaissance de leur travail renforce leur engagement envers la production de contenus précis et bien conçus.
- 3. Promotion d'une Concurrence Plus Équitable :** Lorsque les créateurs de données sont rémunérés équitablement, cela crée un terrain de jeu plus équilibré dans le secteur de l'IA. Les petites entreprises et les indépendants auront un meilleur accès aux ressources nécessaires pour concurrencer les grandes entreprises, qui ont traditionnellement dominé l'espace de l'IA grâce à leurs vastes collections de données.
- 4. Renforcement de l'Éthique et de la Responsabilité :** La rémunération équitable renforce les principes éthiques dans le domaine de l'IA. Elle assure que les bénéfices de l'IA ne sont pas construits sur l'exploitation non rémunérée du travail créatif d'autrui. Cela contribue à un environnement plus responsable et transparent dans le secteur de l'IA.
- 5. Encouragement à la Participation Diversifiée :** En garantissant une rémunération juste, on encourage la participation d'une gamme plus large d'auteurs, y compris ceux de milieux sous-représentés. Cela enrichit l'IA avec une multitude de perspectives et d'expériences, ce qui est crucial pour éviter les biais et promouvoir l'inclusivité dans les modèles d'IA.

6. **Favoriser la Durabilité à Long Terme :** Un modèle de rémunération équitable contribue à la durabilité à long terme de l'écosystème de l'IA. En assurant que les auteurs sont rémunérés pour leurs contributions, on garantit un approvisionnement continu et renouvelé de données de haute qualité, essentielles pour la croissance et l'évolution constantes de l'IA.

En résumé, une rémunération équitable pour les auteurs de données a le potentiel de transformer le secteur de l'IA de manière positive. Elle encourage l'innovation, assure une qualité supérieure des données, promeut une concurrence équitable, renforce l'éthique et la responsabilité, favorise la diversité et soutient la durabilité à long terme. Ces avantages démontrent clairement que la rémunération équitable est non seulement une question de justice, mais aussi un moteur essentiel pour l'avenir de l'IA.

## CADRES RÉGLEMENTAIRES ET LÉGAUX : SOUTENIR LA RÉMUNÉRATION ÉQUITABLE À TRAVERS LA LOI

La mise en place d'un cadre réglementaire et légal robuste est essentielle pour soutenir les modèles de rémunération équitable des auteurs dans le domaine de l'intelligence artificielle (IA) générative. Voici un examen des besoins et des initiatives en cours pour assurer la protection des droits des auteurs :

1. **Adaptation des Lois sur les Droits d'Auteur :** Les lois actuelles sur les droits d'auteur doivent être adaptées pour refléter les défis uniques posés par l'IA. Cela implique de reconnaître les contributions des auteurs de données comme étant éligibles à la protection du droit d'auteur et de définir clairement comment ces droits s'appliquent dans le contexte de l'IA, notamment en ce qui concerne la transformation et la réutilisation des données.
2. **Développement de Réglementations Spécifiques à l'IA :** Il est impératif de développer des réglementations spécifiques qui abordent directement l'utilisation des données dans l'IA. Cela inclut des directives sur la traçabilité des données, la transparence de leur utilisation, et la rémunération équitable des auteurs.
3. **Initiatives pour la Traçabilité et la Transparence :** Des initiatives législatives pour renforcer la traçabilité des données et la transparence de leur utilisation sont nécessaires. Cela pourrait impliquer des exigences pour que les développeurs d'IA documentent l'origine des données et les modalités de leur utilisation.
4. **Contrats de Licence et Rémunération Automatisée :** Les cadres légaux peuvent encourager l'utilisation de contrats de licence clairs et de systèmes de rémunéra-

tion automatisée, comme ceux basés sur la blockchain, pour assurer une rémunération équitable et efficace des auteurs.

5. **Protection Internationale et Collaboration :** Étant donné la nature globale de l'IA, il est crucial d'établir une protection internationale des droits des auteurs. Cela pourrait être réalisé par des traités internationaux ou des accords bilatéraux entre pays.
6. **Initiatives et Propositions de Loi :** Plusieurs pays ont commencé à explorer des législations dans ce domaine. Par exemple, l'Union européenne étudie comment ses directives sur le droit d'auteur peuvent être adaptées pour mieux répondre aux défis de l'IA. Aux États-Unis, des propositions de loi visent à étendre la protection du droit d'auteur aux œuvres générées ou modifiées par l'IA.
7. **Sensibilisation et Débat Public :** Enfin, il est essentiel de promouvoir la sensibilisation et le débat public sur ces questions. Les décideurs politiques, les créateurs de contenu, les entreprises d'IA, et le public doivent collaborer pour façonner un environnement législatif qui soutient une rémunération équitable et l'innovation dans le domaine de l'IA.

En résumé, l'élaboration de cadres réglementaires et légaux adaptés est cruciale pour assurer une rémunération équitable des auteurs dans le secteur de l'IA générative. Ces cadres doivent être souples, prospectifs et en phase avec les progrès technologiques, tout en garantissant la protection et la reconnaissance adéquates des droits des auteurs.

## IMPLICATIONS ÉTHIQUES ET ÉCONOMIQUES DE LA NON-RÉMUNÉRATION DES AUTEURS : UNE VUE D'ENSEMBLE

L'absence de rémunération adéquate pour les auteurs de données utilisées dans les modèles d'intelligence artificielle (IA) générative a des implications profondes, tant sur le plan éthique qu'économique. Ces implications affectent non seulement les auteurs individuels, mais aussi l'industrie de l'IA dans son ensemble et la société en général.

- 1. Éthique de la Rétribution :** Du point de vue éthique, ne pas rémunérer les auteurs pour leur travail constitue une forme d'exploitation. Cela va à l'encontre des principes fondamentaux de justice et d'équité, en utilisant les fruits du travail créatif d'autrui sans compensation. Cette pratique soulève des questions sur le respect des droits de propriété intellectuelle et la valeur que la société attribue à la création artistique et intellectuelle.
- 2. Impact sur la Création de Contenu :** Lorsque les auteurs ne sont pas rémunérés pour leur travail, cela peut décourager la création de contenu de qualité. La non-rémunération envoie un message que leur travail n'est pas valorisé, ce qui peut freiner la motivation et réduire l'incitation à produire des œuvres originales et innovantes. Cela peut, à terme, appauvrir la diversité et la qualité du contenu disponible pour l'entraînement des IA.
- 3. Innovation et Développement Technologique :** L'innovation dans le domaine de l'IA dépend fortement de l'accès à des données diversifiées et de haute qualité. Si les auteurs ne sont pas incités à partager leurs données en raison d'une absence de rémunération, cela pourrait entraver le progrès technologique. Une innovation saine et soutenue nécessite un écosystème où les contributions de chacun sont reconnues et valorisées.
- 4. Justice Économique et Équité :** Sur le plan économique, la non-rémunération des auteurs crée un déséquilibre. Les entreprises qui développent et commercialisent des technologies d'IA peuvent générer d'importants bénéfices en utilisant des données pour lesquelles elles ne paient pas. Cela soulève des préoccupations en matière d'équité économique, car les bénéfices sont accumulés par quelques-uns tandis que les créateurs de contenu ne reçoivent pas leur juste part.
- 5. Effet sur le Marché du Travail et la Répartition des Ressources :** La non-rémunération des auteurs peut également avoir un impact négatif sur le marché du travail global. Elle dévalue le travail créatif et intellectuel, conduisant à une répartition inéquitable des ressources dans l'économie. Cela peut accentuer les inégalités existantes et affaiblir le tissu économique et social.
- 6. Considérations Socio-économiques :** Enfin, la non-rémunération des auteurs a des implications socio-économiques plus larges. Elle peut décourager les talents dans des domaines créatifs et intellectuels, particulièrement chez les jeunes générations, et contribuer à une société où l'innovation créative n'est pas suffisamment soutenue ou valorisée.

En conclusion, les implications éthiques et économiques de la non-rémunération des auteurs soulignent la nécessité de développer des modèles de rémunération équitables et durables dans le domaine de l'IA générative. Cela ne concerne pas seulement les auteurs individuels, mais touche également à la santé globale de l'industrie de l'IA et à l'équité au sein de la société.

# BIBLIOGRAPHIE

- 1 Ouverture de la 5<sup>e</sup> édition de l'Assemblée des idées : « L'intelligence artificielle : sommes-nous prêts ? », Présidence de l'Assemblée Nationale, Mardi 30 mai 2023  
<https://presidence.assemblee-nationale.fr/activites/discours/33094>
- 2 Nos « données personnelles » nous appartiennent : monétisons-les ! Bruno Bonnell, Laurence Parisot, Jaron Lanier et al., Le Monde 5 février 2018  
[https://www.lemonde.fr/idees/article/2018/02/05/nos-donnees-nous-appartiennent-monetisons-les\\_5251774\\_3232.html](https://www.lemonde.fr/idees/article/2018/02/05/nos-donnees-nous-appartiennent-monetisons-les_5251774_3232.html)
- Vers un droit de propriété des données personnelles. Eric Bonnet, Alain Bensoussan. Lexing, 30 mai 2022.  
<https://www.alain-bensoussan.com/avocats/droit-de-proprieté-des-donnees-personnelles/2022/05/30/>
- 3 Interview de Jean-Noël Barrot sur BFM, 12 avril 2023  
[https://www.bfmtv.com/economie/replay-emissions/good-evening-business/intelligence-artificielle-comment-gerer-les-droits-d-auteurs\\_VN-202304120694.html](https://www.bfmtv.com/economie/replay-emissions/good-evening-business/intelligence-artificielle-comment-gerer-les-droits-d-auteurs_VN-202304120694.html)
- 4 La régulation du numérique : Chine, États-Unis, France. Aifang Ma, Fondapol, Septembre 2023.  
<https://www.fondapol.org/etude/la-regulation-du-numerique-chine-etats-unis-france>
- 5 IA générative et droits d'auteur : « La culture artificielle ne doit pas suivre le triste chemin de la malbouffe ». Vincent Lorphelin, Le Monde des 17 et 18 septembre 2023  
[https://www.lemonde.fr/idees/article/2023/09/15/ia-generative-et-droits-d-auteur-la-culture-artificielle-ne-doit-pas-suivre-le-triste-chemin-de-la-malbouffe\\_6189539\\_3232.html](https://www.lemonde.fr/idees/article/2023/09/15/ia-generative-et-droits-d-auteur-la-culture-artificielle-ne-doit-pas-suivre-le-triste-chemin-de-la-malbouffe_6189539_3232.html)
- 6 Sacem, SACD, ADAGP, Adami + 70 organismes. Tribune : Construisons dès aujourd'hui une Intelligence Artificielle de rang mondial respectueuse de la propriété littéraire et artistique.  
[https://www.lemonde.fr/idees/article/2023/09/29/intelligence-artificielle-construisons-des-aujourd-hui-une-ia-de-rang-mondial-respectueuse-de-la-proprieté-litteraire-et-artistique\\_6191655\\_3232.html](https://www.lemonde.fr/idees/article/2023/09/29/intelligence-artificielle-construisons-des-aujourd-hui-une-ia-de-rang-mondial-respectueuse-de-la-proprieté-litteraire-et-artistique_6191655_3232.html)
- 7 <https://composeralliance.org/media/1414-authors-performers-and-other-creative-workers-organisations-joint.pdf>
- 8 <https://authorsguild.org/app/uploads/2023/09/Authors-Guild-OpenAI-Class-Action-Complaint-Sep-2023.pdf>
- 9 <https://s3.documentcloud.org/documents/23963237/authors-v-openai.pdf>
- 10 'New York Times' considers legal action against OpenAI as copyright tensions swirl. Bobby Allyn. NPR, August 16, 2023.  
<https://www.npr.org/2023/08/16/1194202562/new-york-times-considers-legal-action-against-openai-as-copyright-tensions-swirl>
- Gilles Babinet : « Le NY Times va porter plainte contre OpenAI pour violation du copyright. Ce qu'on découvre c'est que cette dernière a massivement crawlé le Web, avalant 250 sites sans rien demander à personne et avec la bénédiction et le soutien de Bing. »  
<https://twitter.com/babgi/status/1692460260215136391>
- 11 <https://drive.google.com/file/d/1jONWdRbwbS50hd1-x4fDvSyARJMCgRTY/view?pli=1>
- 12 IA : les médias français s'organisent face à la collecte de données par les robots. Stéphane Loignon, Les Echos, 28 août 2023  
<https://www.lesechos.fr/tech-medias/medias/ia-les-medias-francais-sorganisent-face-a-la-collecte-de-donnees-par-les-robots-1973079>
- 13 Pour une intelligence artificielle au service de la création, des auteurs et respectueuse de leurs droits. SACD, 31 août 2023  
<https://www.sacd.fr/fr/pour-une-intelligence-artificielle-au-service-de-la-creation-des-auteurs-et-respectueuse-de-leurs>
- 14 Elon Musk sues four unknown individuals for scraping Twitter's data. Emma Roth, The Verge, July 13, 2023.  
<https://www.theverge.com/2023/7/13/23794163/elon-musk-lawsuit-data-scraping-twitter-x-corp>  
<https://archive.vn/1M0hH>

- 15 Les créateurs et artistes interprètes du monde entier réclament le respect de leurs droits face à la prolifération de l'IA. Cisac, Aepo-Artis + 6 autres organismes, 20 Juillet 2023.  
<https://www.cisac.org/fr/Actus-Media/articles/les-createurs-et-artistes-interpretes-du-monde-entier-reclament-le-respect-de>
- 16 <https://lmlitigation.com/pdf/03416/silverman-openai-complaint.pdf>
- 17 [https://www.sagaftra.org/files/sa\\_documents/SAG-AFTRA%20AI%20Letter.pdf](https://www.sagaftra.org/files/sa_documents/SAG-AFTRA%20AI%20Letter.pdf)  
<https://www.adami.fr/wp-content/uploads/2023/07/CP-Greve-des-acteurs-et-actrices-EtatsUnis-France-meme.pdf>  
<https://www.aepo-artis.org/open-letter-on-a-i-from-creators-policy-makers/>
- 18 <https://fingfx.thomsonreuters.com/gfx/legaldocs/myvmodloqvr/GOOGLE%20AI%20LAWSUIT%20complaint.pdf>
- 19 <https://lmlitigation.com/pdf/03223/tremblay-openai-complaint.pdf>
- 20 <https://authorsguild.org/advocacy/artificial-intelligence/faq/>
- 21 [https://www.unitedvoiceartists.com/wp-content/uploads/2023/05/230524\\_UVA\\_Manifesto\\_FR\\_4\\_Alphabetical.pdf](https://www.unitedvoiceartists.com/wp-content/uploads/2023/05/230524_UVA_Manifesto_FR_4_Alphabetical.pdf)
- 22 'Plagiarism machines': Hollywood writers and studios battle over the future of AI. Dawn Chmielewski and Lisa Richwine, Reuters, 3 mai 2023  
<https://www.reuters.com/technology/plagiarism-machines-hollywood-writers-studios-battle-over-future-ai-2023-05-03/>
- 23 Reddit Wants to Get Paid for Helping to Teach Big A.I. Systems. Mike Isaac, NY Times, 18 avril 2023  
<https://www.nytimes.com/2023/04/18/technology/reddit-ai-openai-google.html>
- 24 <https://europeanwriterscouncil.eu/call-for-ethical-ai-regulation/>
- 25 AFPE. "No la cagues": la campaña que defiende a los fotógrafos ante el auge de la Inteligencia Artificial. 21 avril 2023  
<https://prnoticias.com/2023/04/21/no-la-cagues-la-campana-que-defiende-a-los-fotografos-ante-el-auge-de-la-inteligencia-artificial/>
- 26 <https://www.humanartistrycampaign.com/>
- 27 <https://docs.justia.com/cases/federal/district-courts/california/candce/3:2023cv00201/407208/67>
- 28 <https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrlkwmvne/GETTY%20IMAGES%20AI%20LAWSUIT%20complaint.pdf>
- 29 [https://githubcopilotlitigation.com/pdf/06823/1-0-github\\_complaint.pdf](https://githubcopilotlitigation.com/pdf/06823/1-0-github_complaint.pdf)
- 30 Sam Altman, PDG d'OpenAI : « Il faut trouver le bon équilibre entre régulation et innovation ». Alexandre Piquard, Le Monde, 26 Mai 2023.  
[https://www.lemonde.fr/economie/article/2023/05/26/sam-altman-pdg-d-openai-il-faut-trouver-le-bon-equilibre-entre-regulation-et-innovation\\_6175034\\_3234.html](https://www.lemonde.fr/economie/article/2023/05/26/sam-altman-pdg-d-openai-il-faut-trouver-le-bon-equilibre-entre-regulation-et-innovation_6175034_3234.html)
- 31 United States District Court, Northern District of California, Sarah Andersen et al v. Stability AI Ltd et al, Order on Motions to Dismiss and Strike, 30 octobre 2023
- 32 Potential Supreme Court clash looms over copyright issues in generative AI training data. Sharon Goldman, VentureBeat, 6 septembre 2023.  
<https://venturebeat.com/ai/potential-supreme-court-clash-looms-over-copyright-issues-in-generative-ai-training-data>
- 33 U.S. Copyright Office Fair Use Index. Copyright Office  
<https://www.copyright.gov/fair-use>
- 34 [https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236\\_FR.html](https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_FR.html)  
[https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html)
- 35 Deux auteurs compositeurs contestent devant la justice les relevés d'écoutes de la sacem dans les discothèques. Roland-Pierre Paringaud, Le Monde 13 décembre 1996.  
[https://www.lemonde.fr/archives/article/1996/12/13/deux-auteurs-compositeurs-contestent-devant-la-justice-les-relevés-d-écoutes-de-la-sacem-dans-les-discothèques\\_3740426\\_1819218.html](https://www.lemonde.fr/archives/article/1996/12/13/deux-auteurs-compositeurs-contestent-devant-la-justice-les-relevés-d-écoutes-de-la-sacem-dans-les-discothèques_3740426_1819218.html)

36 Règles de répartition. SACEM 2023

<https://createurs-editeurs.sacem.fr/brochures-documents/regles-de-repartition-version-applicable-partir-du-4-janvier-2023>

Décision réglementaire du 9 septembre 1987. Ministère de la Culture et de la Communication

<https://www.spre.fr/wp-content/uploads/2021/05/Decision-reglementaire-du-9-septembre-1987-sur-la-remuneration-equitable.pdf>

37 Commission permanente de contrôle des sociétés de perception et de répartition des droits, Rapport 2012. Rapporteur M. Bernard Menasseyre.

<https://www.culture.gouv.fr/Media/Documentation/Rapports/Rapport-2012-de-la-Commission-permanente-de-contrôle-des-sociétés-de-perception-et-de-repartition-des-droits>

38 Droit d'auteur, comment fonctionne la répartition. Sacem

<https://createurs-editeurs.sacem.fr/sacem-et-moi/repartition-droits-auteur>

39 Shutterstock signs 6-year training data agreement with OpenAI. Victor Dey, VentureBeat, 13 juillet 2023

<https://venturebeat.com/ai/shutterstock-signs-6-year-training-data-agreement-with-openai>

AI-generated Content on Shutterstock: Contributor FAQ. Shutterstock, mise à jour du 8 septembre 2023

[https://support.submit.shutterstock.com/s/article/Shutterstock-ai-and-Computer-Vision-Contributor-FAQ?language=en\\_US](https://support.submit.shutterstock.com/s/article/Shutterstock-ai-and-Computer-Vision-Contributor-FAQ?language=en_US)

Shutterstock Data Licensing and the Contributor Fund. Updated June 16, 2023.

[https://support.submit.shutterstock.com/s/article/Shutterstock-Data-Licensing-and-the-Contributor-Fund?language=en\\_US](https://support.submit.shutterstock.com/s/article/Shutterstock-Data-Licensing-and-the-Contributor-Fund?language=en_US)

"AI trainers pay almost nothing per image but Shutterstock has a contributors' fund that bumps up royalties to an average rate, so for the moment the arrangement is gross margin neutral. Photographers are still boned, of course. Not only does robot competition cannibalise their marketplace royalties, the one-off compensation payments Shutterstock offers are distributed pro rata to library size rather than sales".

Shutterstock has found an unlikely ally in AI. <https://www.ft.com/content/faa4b3ca-d3c7-44a8-aafe-f84ceec785d4>

Shutterstock will start selling AI-generated stock imagery with help from OpenAI. James Vincent, The Verge, 25 octobre 2022.

<https://www.theverge.com/2022/10/25/23422359/shutterstock-ai-generated-art-openai-dall-e-partnership-contributors-fund-reimbursement>

OpenAI, Questions for the Record. U.S Senate, 22 Juin 2023

[https://www.judiciary.senate.gov/imo/media/doc/2023-05-16\\_-\\_qfr\\_responses\\_-\\_altman.pdf](https://www.judiciary.senate.gov/imo/media/doc/2023-05-16_-_qfr_responses_-_altman.pdf)

40 AP, Open AI agree to share select news content and technology in new collaboration. Associated Press, 13 juillet 2023

<https://www.ap.org/press-releases/2023/ap-open-ai-agree-to-share-select-news-content-and-technology-in-new-collaboration>

41 FAQ sur Firefly pour les contributeurs Adobe Stock. Mise à jour du 18 octobre 2023. Adobe

<https://helpx.adobe.com/fr/stock/contributor/help/firefly-faq-for-adobe-stock-contributors.html>

42 ChatGPT : « Le risque de manipulation de masse va exacerber la bataille de l'image entre géants du Web ». Vincent Lorphelin, Le Monde, 19 janvier 2023

43 Apple Music Insights: Royalties. Apple, 2 juin 2021

<https://artists.apple.com/support/1124-apple-music-insights-royalty-rate>

44 "Eligible writers who register with the CMO or a participating organization would receive a distribution based on algorithms that take into account the number of works published, the length of those works, and any available sales data"

FAQs on the Authors Guild's Positions and Advocacy Around Generative AI, Authors Guild, 30 juin 2023

<https://authorsguild.org/advocacy/artificial-intelligence/faq/>

Authors call for AI companies to stop using their work without consent. Lucy Knight, The Guardian, 20 Juillet 2023.

<https://www.theguardian.com/books/2023/jul/20/authors-call-for-ai-companies-to-stop-using-their-work-without-consent>

45 FAQs on the Authors Guild's Positions and Advocacy Around Generative AI, Authors Guild, 30 juin 2023

<https://authorsguild.org/advocacy/artificial-intelligence/faq/>

AI learned from their work. Now they want compensation. Washington Post, 16 juillet 2023.

<https://www.washingtonpost.com/technology/2023/07/16/ai-programs-training-lawsuits-fair-use>

46 "Plaintiffs and the Classes should be personally and directly compensated for the fair market value of their contributions to the LLMs on which Bard was built, in an amount to be determined by expert testimony"

Class Action v. Google, Northern District of California, 11 Juillet 2023

<https://fingfx.thomsonreuters.com/gfx/legaldocs/mymodloqvr/GOOGLE%20AI%20LAWSUIT%20complaint.pdf>

47 Sacem, SACD, ADAGP, Adami + 70 organismes. Tribune : Construisons dès aujourd'hui une Intelligence Artificielle de rang mondial respectueuse de la propriété littéraire et artistique.

[https://www.lemonde.fr/idees/article/2023/09/29/intelligence-artificielle-construisons-des-aujourd-hui-une-ia-de-rang-mondial-respectueuse-de-la-propriete-litteraire-et-artistique\\_6191655\\_3232.html](https://www.lemonde.fr/idees/article/2023/09/29/intelligence-artificielle-construisons-des-aujourd-hui-une-ia-de-rang-mondial-respectueuse-de-la-propriete-litteraire-et-artistique_6191655_3232.html)

48 Authors', performers' and other creative workers'organisations joint statement on artificial intelligence and the draft act. United Voice artists + 12 organisations, 26 septembre 2023

<https://composeralliance.org/media/1414-authors-performers-and-other-creative-workers-organisations-joint.pdf>

49 ChatGPT-maker OpenAI signs deal with AP to license news stories. Matt O'Brien, Associated Press, 13 Juillet 2023.

<https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a>

50 Construisons dès aujourd'hui une Intelligence Artificielle de rang mondial respectueuse de la propriété littéraire et artistique. Collectif de 70 organismes professionnels des secteurs de la création et des industries culturelles.

<https://www.sne.fr/actu/tribune-construisons-des-aujourd'hui-une-intelligence-artificielle-de-rang-mondial-respectueuse-de-la-propriete-litteraire-et-artistique>

51 Shutterstock signs 6-year training data agreement with OpenAI. Victor Dey, VentureBeat, 13 juillet 2023

<https://venturebeat.com/ai/shutterstock-signs-6-year-training-data-agreement-with-openai>

AI-generated Content on Shutterstock: Contributor FAQ. Shutterstock, mise à jour du 8 septembre 2023

[https://support.submit.shutterstock.com/s/article/Shutterstock-ai-and-Computer-Vision-Contributor-FAQ?language=en\\_US](https://support.submit.shutterstock.com/s/article/Shutterstock-ai-and-Computer-Vision-Contributor-FAQ?language=en_US)

Shutterstock Data Licensing and the Contributor Fund. Updated June 16, 2023.

[https://support.submit.shutterstock.com/s/article/Shutterstock-Data-Licensing-and-the-Contributor-Fund?language=en\\_US](https://support.submit.shutterstock.com/s/article/Shutterstock-Data-Licensing-and-the-Contributor-Fund?language=en_US)

"AI trainers pay almost nothing per image but Shutterstock has a contributors' fund that bumps up royalties to an average rate, so for the moment the arrangement is gross margin neutral. Photographers are still boned, of course. Not only does robot competition cannibalise their marketplace royalties, the one-off compensation payments Shutterstock offers are distributed pro rata to library size rather than sales".

Shutterstock has found an unlikely ally in AI.

<https://www.ft.com/content/faa4b3ca-d3c7-44a8-aafe-f84ceec785d4>

Shutterstock will start selling AI-generated stock imagery with help from OpenAI. James Vincent, The Verge, 25 octobre 2022.

<https://www.theverge.com/2022/10/25/23422359/shutterstock-ai-generated-art-openai-dall-e-partnership-contributors-fund-reimbursement>

OpenAI, Questions for the Record. U.S Senate, 22 Juin 2023

[https://www.judiciary.senate.gov/imo/media/doc/2023-05-16\\_-\\_qfr\\_responses\\_-\\_altman.pdf](https://www.judiciary.senate.gov/imo/media/doc/2023-05-16_-_qfr_responses_-_altman.pdf)

52 FAQ sur Firefly pour les contributeurs Adobe Stock. Mise à jour du 18 octobre 2023. Adobe

<https://helpx.adobe.com/fr/stock/contributor/help/firefly-faq-for-adobe-stock-contributors.html>

53 Bria to Offer Its Comprehensive, Responsibly Sourced, Generative AI Visual Content Solution for Enterprises on NVIDIA AI Foundations. Bria, 21 mars 2023, PRnewswire

<https://www.prnewswire.com/news-releases/bria-to-offer-its-comprehensive-responsibly-sourced-generative-ai-visual-content-solution-for-enterprises-on-nvidia-ai-foundations-301777666.html>

54 This startup wants to train art-generating AI strictly on licensed images. Kyle Wiggers, Techcrunch, 13 avril 2023

<https://techcrunch.com/2023/04/13/this-startup-wants-to-train-art-generating-ai-strictly-on-licensed-images/?guccounter=1>

Derrière l'intelligence artificielle, le retour d'utopies technologiques. Alexandre Piquard, Le Monde, 13 juin 2023.

[https://www.lemonde.fr/economie/article/2023/06/13/derriere-l-ia-le-retour-d-utopies-technologiques\\_6177367\\_3234.html](https://www.lemonde.fr/economie/article/2023/06/13/derriere-l-ia-le-retour-d-utopies-technologiques_6177367_3234.html)

55 Content Creator License Agreement. Ascendant Art

<https://ascendantart.ai/content-creator-license-agreement>

Ascendant Art Launches AI Avatar App That Pays Royalties to Artists. Ascendant Art, 9 Mai 2023.

<https://www.prnewswire.com/news-releases/ascendant-art-launches-ai-avatar-app-that-pays-royalties-to-artists-301819874.html>

Ascendant AI créée par Mitch Randall en 2018

<https://www.linkedin.com/in/mitch-randall-6b4bba2>

56 AP, Open AI agree to share select news content and technology in new collaboration. Associated Press, 13 juillet 2023

<https://www.ap.org/press-releases/2023/ap-open-ai-agree-to-share-select-news-content-and-technology-in-new-collaboration>

57 Sacem, SACD, ADAGP, Adami + 70 organismes. Tribune : Construisons dès aujourd'hui une Intelligence Artificielle de rang mondial respectueuse de la propriété littéraire et artistique.

[https://www.lemonde.fr/idees/article/2023/09/29/intelligence-artificielle-construisons-des-aujourd-hui-une-ia-de-rang-mondial-respectueuse-de-la-proprietee-litteraire-et-artistique\\_6191655\\_3232.html](https://www.lemonde.fr/idees/article/2023/09/29/intelligence-artificielle-construisons-des-aujourd-hui-une-ia-de-rang-mondial-respectueuse-de-la-proprietee-litteraire-et-artistique_6191655_3232.html)

AI Act: EU Commission attempts to revive tiered approach shifting to General Purpose AI. Luca Bertuzzi, Euractiv, 20 novembre 2023

<https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-commission-attempts-to-revive-tiered-approach-shifting-to-general-purpose-ai>

58 Tweet d'Arthur Mensch, Mistral IA, 16 novembre 2023

<https://twitter.com/arthurmensch/status/1725076260827566562?s=20>

59 Règlementation de l'IA, l'UE franchit une nouvelle étape après un vote clé au Parlement Européen. Luca Bertuzzi, Euractiv, 11 mai 2023.

<https://www.euractiv.fr/section/economie/news/reglementation-de-lia-lue-franchit-une-nouvelle-etape-apres-un-vote-cle-au-parlement-europeen>

Fabienne Schmitt. L'Europe franchit un nouveau pas dans la régulation de l'intelligence artificielle. Les Echos, 12 mai 2023.

<https://www.lesechos.fr/tech-medias/intelligence-artificielle/leurope-franchit-un-nouveau-pas-dans-la-regulation-de-lintelligence-artificielle-1942509>

60 EU's AI act could kill our company,' says Mistral's Cédric O. Zosia Wanat, Sifted, 27 Octobre 2023.

<https://sifted.eu/articles/eu-ai-act-kill-mistral-cedric-o>

AI Act: Leading MEPs propose initial criteria for classifying foundation models as 'high-impact'. Luca Bertuzzi, Euractiv, 9 novembre 2023

[https://www.euractiv.com/section/artificial-intelligence/news/ai-act-leading-meps-propose-initial-criteria-for-classifying-foundation-models-as-high-impact/?\\_ga=2.172005667.1051799477.1700154198-1728994896.1700154198](https://www.euractiv.com/section/artificial-intelligence/news/ai-act-leading-meps-propose-initial-criteria-for-classifying-foundation-models-as-high-impact/?_ga=2.172005667.1051799477.1700154198-1728994896.1700154198)

AI Act : négociations bloquées à cause de divergences sur les modèles de fondation. Luca Bertuzzi, Euractiv, 13 novembre 2023

<https://www.euractiv.fr/section/intelligence-artificielle/news/ai-act-negociations-bloquees-a-cause-de-divergences-sur-les-modeles-de-fondation>

Urgent – AI Act: Letter of concern on the stopped negotiations. European Writers' Council, 14 novembre 2023.

[https://europeanwriterscouncil.eu/2311ai\\_act\\_negotiationsstop](https://europeanwriterscouncil.eu/2311ai_act_negotiationsstop)

AI Act : Paris, Berlin et Rome veulent « auto-réguler » les modèles de fondation. Luca Bertuzzi, Euractiv, 20 novembre 2023

<https://www.euractiv.fr/section/intelligence-artificielle/news/ai-act-paris-berlin-et-rome-veulent-auto-reguler-les-modeles-de-fondation>

61 The worlds most valuable resource us no longer oil but data. The Economist, 6 Mai 2017.

<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

Data is the new oil...who's getting to own it ? Guy Chazan, Financial Times, 16 novembre 2016.

<https://www.ft.com/content/e548deac-856a-11e6-8897-2359a58ac7a5>

As personal data get trickier to come by, AI is swallowing everything else. The Economist, 31 août 2023.

<https://www.economist.com/podcasts/2023/08/31/as-personal-data-get-trickier-to-come-by-ai-is-swallowing-everything-else>

62 Intelligence artificielle et droit d'auteur Georgie Courtois, Jean-Sébastien Mariez, Jeanne Roussel, Avocate, De Gaulle Fleurance & Associés

<https://www.culture.gouv.fr/content/download/263325/file/Annexe%204-%20DGFLA.pdf>

- 63 Tristan Gaudiaut. Comment les géants de la tech gagnent-ils leurs milliards ? Statista, 3 mai 2022.  
<https://fr.statista.com/infographie/13194/principales-sources-de-revenus-gafam-geants-tech-part-segment-activites-produits-dans-chiffre-affaires>
- 64 Hayat Gazzane. Pétrole : l'Opep devrait engranger plus de 900 milliards de dollars de recettes cette année. Les Echos, 23 août 2022  
<https://www.lesechos.fr/finance-marches/marches-financiers/petrole-lopep-devrait-engranger-plus-de-900-milliards-de-dollars-de-recettes-cette-annee-1783245>
- Pétrole : les recettes des pays de l'Opep bondissent de 50%. Robert Jules, La Tribune, 30 août 2022  
<https://www.latribune.fr/economie/international/petrole-une-manne-a-842-milliards-de-dollars-pour-l-opep-928776.html>
- 65 CNUCED. Nouvelle étude : le commerce mondial du plastique dépasse de 40% les évaluations précédentes. 3 mars 2021  
<https://unctad.org/fr/news/nouvelle-etude-le-commerce-mondial-du-plastique-depasse-de-40-les-evaluations-precedentes>
- La banque UBS a estimé que le marché adressable total de l'IA générative pourrait représenter 1.000 milliards de dollars. MARCHÉ : CHATGPT DÉCLENCHE UNE FLAMBÉE EN BOURSE AUTOUR DE L'INTELLIGENCE ARTIFICIELLE. BFM Bourse, 11 février 2023  
<https://www.tradingsat.com/actualites/marches-financiers/chatgpt-declenche-une-flambee-en-bourse-autour-de-l-intelligence-artificielle-1056251.html>
- 66 CISAC Global Collection Report 2022.  
<https://members.cisac.org/CisacPortal/cisacDownloadFileSearch.do?docId=43279&lang=en>
- 67 Aramco : (i) the royalty rate was reduced to 15% (from 20%) on Brent prices up to \$70 per barrel; (ii) the marginal royalty rate increased to 45% (from 40%) on Brent prices above \$70 per barrel up to \$100 per barrel; and (iii) the marginal royalty rate increased to 80% (from 50%) on Brent prices above \$100 per barrel. (p.118)  
<https://www.aramco.com/-/media/publications/corporate-reports/saudi-aramco-ara-2022-english.pdf?la=en&hash=6BC0409B50ECF-F4A4C743307DF2FF7BDBCEC8B43>
- Aramco : les coûts d'extraction les plus bas du monde. Vincent Collen, les Echos, 10 avril 2019  
<https://www.lesechos.fr/finance-marches/marches-financiers/aramco-les-couts-d'extraction-les-plus-bas-du-monde-1008043>
- U.S Federal lands  
<https://www.taxpayer.net/wp-content/uploads/2020/02/TCS-Royally-Losing-2020.pdf>
- 68 « Les données en tant que telles ont souvent peu de valeur, mais en gagnent quand elles sont contextualisées, croisées avec d'autres. »  
Donner du sens à l'intelligence artificielle : pour une stratégie nationale et européenne. Rapport de Cédric Villani  
[https://www.aiforhumanity.fr/pdfs/9782111457089\\_Rapport\\_Villani\\_accessible.pdf](https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf)
- 69 L'écriture des films et séries en France. CNC, SACD, Avril 2019  
[https://www.sacd.fr/sites/default/files/etude\\_cnc\\_sacd\\_ecriture\\_films\\_et\\_series\\_tv\\_2019\\_04\\_09.pdf](https://www.sacd.fr/sites/default/files/etude_cnc_sacd_ecriture_films_et_series_tv_2019_04_09.pdf)
- 70 9e BAROMÈTRE DES RELATIONS AUTEURS/ÉDITEURS. Scam, SGDL  
[https://www.scam.fr/uploads/2023/03/9e\\_Barometre\\_2023.pdf](https://www.scam.fr/uploads/2023/03/9e_Barometre_2023.pdf)
- 71 Mode d'emploi pour le calcul des droits d'auteur à l'usage des compagnies professionnelles indépendantes. SACD  
<https://www.sacd.fr/fr/mode-emploi-calcul-droits>
- 72 Beyond Markets and States: Polycentric Governance of Complex Economic Systems Prize Lecture, December 8, 2009 by Elinor Ostrom  
[https://www.nobelprize.org/uploads/2018/06/ostrom\\_lecture.pdf](https://www.nobelprize.org/uploads/2018/06/ostrom_lecture.pdf)
- 73 Droits voisins : Une étude démontre que Google devrait rendre CHF 154 millions aux éditeurs suisses. Victoria Marchand, Comin-Mag, 17 mars 2023  
<https://cominmag.ch/droits-voisins-une-etude-demontre-que-google-devrait-rendre-chf-154-millions-aux-editeurs-suisses>
- « Droit voisin » : la France devient le premier pays à transposer la directive européenne. Le Monde, 23 juillet 2019.  
[https://www.lemonde.fr/actualite-medias/article/2019/07/23/droit-voisin-la-loi-ouvre-la-porte-a-une-dure-negociation-entre-medias-et-plates-formes\\_5492480\\_3236.html](https://www.lemonde.fr/actualite-medias/article/2019/07/23/droit-voisin-la-loi-ouvre-la-porte-a-une-dure-negociation-entre-medias-et-plates-formes_5492480_3236.html)

The Sad Truth About How Much Your Facebook Data is Worth on the Dark Web, Maria LaMagna, Marketwatch June 6, 2018  
<https://www.marketwatch.com/story/spooked-by-the-facebook-privacy-violations-this-is-how-much-your-personal-data-is-worth-on-the-dark-web-2018-03-20>

[X – Twitter] The cheapest, Small Package, gives access to 50 million tweets for \$42,000 a month. Higher tiers give researchers or businesses access to larger volumes of tweets—100 million and 200 million tweets respectively—and cost \$125,000 and \$210,000 a month.

Twitter's \$42,000-per-Month API Prices Out Nearly Everyone. CHRIS STOKEL-WALKER, Wired, 10 mars 2023  
<https://www.wired.com/story/twitter-data-api-prices-out-nearly-everyone>

There are a variety of methods for estimating [Wikipedia's] value in terms of its market value, its replacement cost, and the value it creates for its users. These methods suggest a valuation in the tens of billions of dollars, a one-time replacement cost of \$6.6 billion with an annual updating cost of \$630 million, and consumer benefit in the hundreds of billions of dollars.

Wikipedia's Economic Value, Jonathan Band and Jonathan Gerafi, 7 octobre 2013.  
<https://ssrn.com/abstract=2338563> or <http://dx.doi.org/10.2139/ssrn.2338563>

Property, Privacy, and Personal Data. Paul M. Schwartz HARV. L. REV. 2056. May, 2004

The Economics of Privacy. Alessandro Acquisti et al. J. OF ECON. LITERATURE 442, 444. Mar. 8, 2016

74 Streaming Drives Global Music Industry Resurgence. Felix Richter, Statista, 6 octobre 2023  
<https://www.statista.com/chart/4713/global-recorded-music-industry-revenues>

Auteurs et compositeurs, parents pauvres du streaming. Nicole Vulser, le Monde, 28 septembre 2022  
[https://www.lemonde.fr/economie/article/2022/09/28/auteurs-et-compositeurs-parents-pauvres-du-streaming\\_6143495\\_3234.html](https://www.lemonde.fr/economie/article/2022/09/28/auteurs-et-compositeurs-parents-pauvres-du-streaming_6143495_3234.html)

Study on the place and role of authors and composers in the European music streaming market. Emmanuel Legrand of Legrand Network, commissioned by GESAC – European Grouping of Societies of Authors and Composer, Septembre 2002  
<https://authorsocieties.eu/content/uploads/2022/09/music-streaming-study-28-9-2022.pdf>

Apple Music Insights: Royalties. Apple, 2 juin 2021  
<https://artists.apple.com/support/1124-apple-music-insights-royalty-rate>

Droits d'auteur, comment fonctionne la répartition. Sacem  
<https://createurs-editeurs.sacem.fr/sacem-et-moi/repartition-droits-auteur>

75 [en 1960] le marché pétrolier est organisé et structuré par l'action des compagnies pétrolières anglo-saxonnes intégrées, familièrement appelées les « sept sœurs ». Elles contrôlent, dans l'après-guerre, 80 % des gisements de pétrole..Alors que le marché de l'or noir était depuis longtemps sous l'emprise des grandes multinationales américaines et européennes, les grands pays exportateurs vont en quelques années récupérer la maîtrise de l'extraction et de la politique des prix.

Après le choc pétrolier d'octobre 1973, l'économie mondiale à l'épreuve du pétrole cher. Céline Antonin, Revue internationale et stratégique 2013/3 (n° 91), pages 139 à 149  
<https://www.cairn.info/revue-internationale-et-strategique-2013-3-page-139.htm>

Le Secret des sept sœurs. Martine Delahaye, Le Monde, 01 octobre 2011  
[https://www.lemonde.fr/vous/article/2011/10/01/le-secret-des-sept-soeurs\\_1580999\\_3238.html](https://www.lemonde.fr/vous/article/2011/10/01/le-secret-des-sept-soeurs_1580999_3238.html)

76 « Une puissance renforcée par un facteur stratégique décisif, qui fragilisait au contraire la première alliance délicate des pays exportateurs : une abondance sans égale de sources de brut disponibles.» Cartel contre cartel : la pénible émergence de l'Opep. Matthieu Auzanneau, Or noir (2015), pages 385 à 400  
<https://www.cairn.info/or-noir--9782707190628-page-385.htm>

« Ces concessions coûtaient aux «sept sœurs» 50% des revenus pétroliers, calculés sur un prix dit «posté» du baril, fixé arbitrairement par le cartel des «supermajors». Les «sept sœurs» décident alors d'entreprendre, pour la première fois, des négociations avec un interlocuteur unique, l'Opep, comptant sur des dissensions internes pour négocier en position de force. Sauf que les 10 membres du cartel des pays exportateurs réunis à Téhéran en janvier 1971 font cause commune et parviennent à un accord qui prévoit une augmentation graduelle du prix du pétrole sur une période de cinq ans, avant que les compagnies concèdent une augmentation du prix plus importante encore quelques mois plus tard à Tripoli ». OPEP ou 60 ans de guerre des cartels pétroliers. Tradingsat, mardi 15 septembre 2020  
<https://www.tradingsat.com/actualites/marches/l-opep-ou-60-ans-de-guerre-des-cartels-petroliers-930413.html>

- 77 Pétrole : deux dollars en 1970, 130 en 2008: quatre décennies d'ascension du baril. Les Echos, 21 mai 2008  
<https://www.lesechos.fr/2008/05/petrole-deux-dollars-en-1970-130-en-2008-quatre-decennies-dascension-du-baril-chronologie-510494>
- 78 Bruno Racine, L'auteur et l'acte de création, Janvier 2020  
<https://www.culture.gouv.fr/Espace-documentation/Rapports/L-auteur-et-l-acte-de-creation>
- 79 Directive du Parlement Européen et du Conseil sur le droit d'auteur dans le marché unique numérique. 14 septembre 2016  
<https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52016PC0593&from=EN>
- 80 Bruno Racine, L'auteur et l'acte de création, Janvier 2020  
<https://www.culture.gouv.fr/Espace-documentation/Rapports/L-auteur-et-l-acte-de-creation>
- 81 Proposition de loi visant à encadrer l'intelligence artificielle par le droit d'auteur. Guillaume VUILLETET, Assemblée Nationale, 12 septembre 2023  
[https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630\\_proposition-loi](https://www.assemblee-nationale.fr/dyn/16/textes/l16b1630_proposition-loi)
- 82 Protection des droits des auteurs et artistes-interprètes : transposition de la directive sur 17 avril 2019. Johanna Bacouelle, Village Justice, 27 mai 2021  
<https://www.village-justice.com/articles/protection-des-droits-des-auteurs-des-artistes-interpretes-transposition,39263.html>
- AI learned from their work. Now they want compensation. Gerrit De Vynck, The Washington Post, 16 juillet 2023.  
<https://www.washingtonpost.com/technology/2023/07/16/ai-programs-training-lawsuits-fair-use>
- Generative AI should pay human artists for training. Financial Times  
<https://www.ft.com/content/c42189e0-4069-4e17-8dc0-72544dc1d51b>
- 83 The Authors Guild's Positions and Advocacy Around Generative AI. The Authors Guild, Juin 2023.  
<https://authorsguild.org/advocacy/artificial-intelligence/faq>
- 84 Les Echos, 19 Juin 2023.  
<https://www.lesechos.fr/tech-medias/intelligence-artificielle/les-internautes-font-massivement-confiance-aux-ia-generatives-1953706>
- 85 Just 42% of Polled Americans Think Artists Should Be Compensated When AI Uses Their Music. Ashley King, Digital Music News, 11 mai 2023  
<https://www.digitalmusicnews.com/2023/05/11/just-42-percent-of-polled-americans-think-artists-should-be-compensated-when-ai-uses-their-music>
- 86 Generative AI should pay human artists for training.  
<https://www.ft.com/content/c42189e0-4069-4e17-8dc0-72544dc1d51b>
- 87 Martine Robert, Filière musicale : le rapport Bargeton préconise une taxe sur le streaming gratuit et payant, Les Echos, 20 avril 2023  
<https://www.lesechos.fr/industrie-services/services-conseils/filiere-musicale-le-rapport-bargeton-preconise-une-taxe-sur-le-streaming-gratuit-et-payant-1936692>
- 88 Yuval Noah Harari argues that AI has hacked the operating system of human civilization. The Economist, 28 avril 2023.  
[https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation?utm\\_campaign=a.coronavirus-special-edition&utm\\_medium=email.internal-newsletter.np&utm\\_source=salesforce-marketing-cloud&utm\\_term=5/20/2023&utm\\_id=1606560](https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation?utm_campaign=a.coronavirus-special-edition&utm_medium=email.internal-newsletter.np&utm_source=salesforce-marketing-cloud&utm_term=5/20/2023&utm_id=1606560)
- 89 L'intelligence artificielle serait aussi dangereuse que « les pandémies ou la guerre nucléaire », selon des leaders du secteur. Alexandre Piquard, Le Monde, 30 mai 2023  
[https://www.lemonde.fr/economie/article/2023/05/30/l-intelligence-artificielle-serait-aussi-dangereuse-que-les-pandemies-ou-la-guerre-nucleaire-selon-des-leaders-du-secteur\\_6175471\\_3234.html](https://www.lemonde.fr/economie/article/2023/05/30/l-intelligence-artificielle-serait-aussi-dangereuse-que-les-pandemies-ou-la-guerre-nucleaire-selon-des-leaders-du-secteur_6175471_3234.html)
- La lettre ouverte choc qui met en garde l'humanité contre le danger de l'intelligence artificielle. Les Echos, 29 mars 2023.  
<https://www.lesechos.fr/tech-medias/hightech/la-lettre-ouverte-choc-qui-met-en-garde-lhumanite-contre-le-danger-de-lintelligence-artificielle-1920390>
- 90 L'intelligence artificielle, nouvelle venue parmi les risques majeurs. Marc Angrand, Le Monde, 29 Octobre 2023.  
[https://www.lemonde.fr/economie/article/2023/10/29/l-intelligence-artificielle-nouvelle-venue-parmi-les-risques-majeurs\\_6197221\\_3234.html](https://www.lemonde.fr/economie/article/2023/10/29/l-intelligence-artificielle-nouvelle-venue-parmi-les-risques-majeurs_6197221_3234.html)

- 91 ChatGPT : « Le risque de manipulation de masse va exacerber la bataille de l'image entre géants du Web ». Vincent Lorphelin, Le Monde, 19 Janvier 2023  
[https://www.lemonde.fr/idees/article/2023/01/19/chatgpt-le-risque-de-manipulation-de-masse-va-exacerber-la-bataille-de-l-image-entre-geants-du-Web\\_6158563\\_3232.html](https://www.lemonde.fr/idees/article/2023/01/19/chatgpt-le-risque-de-manipulation-de-masse-va-exacerber-la-bataille-de-l-image-entre-geants-du-Web_6158563_3232.html)
- 92 Le Royaume-Uni organise le premier sommet mondial sur les risques associés à l'intelligence artificielle. Cécile Ducourtieux, Le Monde, 30 octobre 2023.  
[https://www.lemonde.fr/international/article/2023/10/30/le-royaume-uni-organise-le-premier-sommet-mondial-sur-les-risques-associes-a-l-intelligence-artificielle\\_6197300\\_3210.html](https://www.lemonde.fr/international/article/2023/10/30/le-royaume-uni-organise-le-premier-sommet-mondial-sur-les-risques-associes-a-l-intelligence-artificielle_6197300_3210.html)
- 93 Elon Musk: I'm A Speciesist And Pro-Human, Google's Larry Page Wants To Become A «Digital God». RealClear Politics, 17 avril 2023.  
[https://www.realclearpolitics.com/video/2023/04/17/elon\\_musk\\_im\\_a\\_specist\\_and\\_pro-human\\_googles\\_larry\\_page\\_wants\\_to\\_become\\_a\\_digital\\_god.html](https://www.realclearpolitics.com/video/2023/04/17/elon_musk_im_a_specist_and_pro-human_googles_larry_page_wants_to_become_a_digital_god.html)
- 94 <https://twitter.com/elonmusk/status/1607590239874211847>
- 95 Elon Musk says we need to regulate AI before it becomes a danger to humanity. James Vincent, The Verge, 17 Juillet 2017  
<https://www.theverge.com/2017/7/17/15980954/elon-musk-ai-regulation-existential-threat>
- 96 Alexandre Piquard. Le créateur de ChatGPT passe le test de l'audition au Congrès américain. Le Monde, 16 mai 2023  
[https://www.lemonde.fr/economie/article/2023/05/16/chatgpt-son-createur-passe-le-test-de-l-audition-au-congres-americain\\_6173631\\_3234.html](https://www.lemonde.fr/economie/article/2023/05/16/chatgpt-son-createur-passe-le-test-de-l-audition-au-congres-americain_6173631_3234.html)
- 97 Alexandre Piquard. Derrière l'intelligence artificielle, le retour d'utopies technologiques. Le Monde, 13 juin 2023  
[https://www.lemonde.fr/economie/article/2023/06/13/derriere-l-ia-le-retour-d-utopies-technologiques\\_6177367\\_3234.html](https://www.lemonde.fr/economie/article/2023/06/13/derriere-l-ia-le-retour-d-utopies-technologiques_6177367_3234.html)
- 98 Techlash : " growing public animosity towards large Silicon Valley platform technology companies and their Chinese equivalents". Kari Paul, The Guardian, 28 décembre 2019  
<https://www.theguardian.com/technology/2019/dec/28/tech-industry-year-in-review-facebook-google-amazon>
- Americans have become much less positive about tech companies' impact on the U.S. CARROLL DOHERTY AND JOCELYN KILEY, Pew Research Center, 29 Juillet 2019  
<https://www.pewresearch.org/short-reads/2019/07/29/americans-have-become-much-less-positive-about-tech-companies-impact-on-the-u-s>
- 99 The spread of techlash. Ipsos, 17 mars 2019  
<https://www.ipsos.com/en/spread-techlash>
- 100 Twitter se retire du code européen contre la désinformation en ligne. Le Monde, 27 mai 2023  
[https://www.lemonde.fr/pixels/article/2023/05/27/twitter-se-retire-du-code-europeen-contre-la-desinformation-en-ligne\\_6175078\\_4408996.html](https://www.lemonde.fr/pixels/article/2023/05/27/twitter-se-retire-du-code-europeen-contre-la-desinformation-en-ligne_6175078_4408996.html)
- 101 GAFA : « Il faut prendre la Bastille numérique ». Interview de Sébastien Soriano, président de l'Arcep, réalisée par Marc Chevallier et Justin Delépine pour Alternative Économique, 25 septembre 2018  
<https://www.arcep.fr/actualites/les-prises-de-parole/detail/n/gafa-il-faut-prendre-la-bastille-numerique.html>
- Les géants du numérique, des nains en termes d'emploi. Alexandre Mirlicourtois, Xerfi Canal, 13 avril 2017  
[https://www.xerficanal.com/economie/emission/Alexandre-Mirlicourtois\\_Les-GAFA-riches-a-milliards-pauvres-en-emplois\\_3744644.html](https://www.xerficanal.com/economie/emission/Alexandre-Mirlicourtois_Les-GAFA-riches-a-milliards-pauvres-en-emplois_3744644.html)
- 102 Contre la surveillance en ligne, des internautes « empoisonnent » leurs données personnelles. Aurélien Defer, le Monde, 29 avril 2022  
[https://www.lemonde.fr/pixels/article/2022/04/29/contre-la-surveillance-en-ligne-des-internautes-empoisonnent-leurs-donnees-personnelles\\_6124107\\_4408996.html](https://www.lemonde.fr/pixels/article/2022/04/29/contre-la-surveillance-en-ligne-des-internautes-empoisonnent-leurs-donnees-personnelles_6124107_4408996.html)
- 103 Intelligence artificielle : le Parlement européen devrait convoquer immédiatement Elon Musk et l'auditionner à Bruxelles. Vincent Lorphelin, Le Monde, 15 avril 2023  
[https://www.lemonde.fr/idees/article/2023/04/15/intelligence-artificielle-le-parlement-europeen-devrait-convoquer-immEDIATEMENT-elon-musk-et-l-auditionner-a-bruxelles\\_6169657\\_3232.html](https://www.lemonde.fr/idees/article/2023/04/15/intelligence-artificielle-le-parlement-europeen-devrait-convoquer-immEDIATEMENT-elon-musk-et-l-auditionner-a-bruxelles_6169657_3232.html)

- 104 Yoshua Bengio, chercheur : « Aujourd'hui, l'intelligence artificielle, c'est le Far West ! Nous devons ralentir et réguler ». Claire Legros, Le Monde, 28 avril 2023  
[https://www.lemonde.fr/idees/article/2023/04/28/yoshua-bengio-chercheur-aujourd-hui-l-intelligence-artificielle-c-est-le-far-west-nous-devons-ralentir-et-reguler\\_6171336\\_3232.html](https://www.lemonde.fr/idees/article/2023/04/28/yoshua-bengio-chercheur-aujourd-hui-l-intelligence-artificielle-c-est-le-far-west-nous-devons-ralentir-et-reguler_6171336_3232.html)
- 105 Geoffrey Hinton, pionnier de l'IA, quitte Google et dit « regretter » son invention. Le Monde, 2 Mai 2023  
[https://www.lemonde.fr/pixels/article/2023/05/02/geoffrey-hinton-pionnier-de-l-ia-quitte-google-et-dit-regretter-son-invention\\_6171773\\_4408996.html](https://www.lemonde.fr/pixels/article/2023/05/02/geoffrey-hinton-pionnier-de-l-ia-quitte-google-et-dit-regretter-son-invention_6171773_4408996.html)
- 106 Deezer et Universal repensent le modèle de rémunération du streaming musical. Stéphane Loignon, Les Echos, 6 septembre 2023  
<https://www.lesechos.fr/tech-medias/medias/deezer-et-universal-repensent-le-modele-de-remuneration-du-streaming-musical-1975865>
- 107 « Approximately three AI images are added for every two conventional ones » Bryce Elder. Shutterstock has found an unlikely ally in AI. Financial Times, 23 Mars 2023  
<https://www.ft.com/content/faa4b3ca-d3c7-44a8-aafe-f84ceec785d4>
- “Over the last two years alone 90 percent of the data in the world was generated.” How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read. Bernard Marr, Forbes, 21 Mai 2018  
<https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=3651c42860ba>
- 108 L'intelligence artificielle peut-elle s'effondrer sur elle-même ? Corentin Benoit-Gonin, Le Monde, 10 septembre 2023  
[https://www.lemonde.fr/pixels/article/2023/09/10/l-intelligence-artificielle-peut-elle-s-effondrer-sur-elle-meme\\_6188777\\_4408996.html](https://www.lemonde.fr/pixels/article/2023/09/10/l-intelligence-artificielle-peut-elle-s-effondrer-sur-elle-meme_6188777_4408996.html)
- Emmanuel Macron. Discours du Président à Vivotech : « Pour cela, il faut aussi créer des bases de données en français. Sinon nous utiliserons des modèles qui auront des biais hérités des Anglo-Saxons », a souligné le président, voyant dans l'utilisation de la langue française dans l'IA un enjeu de souveraineté et de soft power culturel. Pour permettre l'entraînement des IA, l'Etat devra donc « ouvrir » ses bases de données, a-t-il dit. « Nous croyons à l'open source », a-t-il ajouté. IA : Emmanuel Macron veut créer des concurrents français des modèles d'OpenAI ou Google. Alexandre Piquard et Vincent Fagot, Le Monde, 14 Juin 2023  
[https://www.lemonde.fr/economie/article/2023/06/14/ia-emmanuel-macron-veut-creer-des-concurrents-francais-des-modeles-d-openai-ou-google\\_6177679\\_3234.html](https://www.lemonde.fr/economie/article/2023/06/14/ia-emmanuel-macron-veut-creer-des-concurrents-francais-des-modeles-d-openai-ou-google_6177679_3234.html)
- 109 « Beaumarchais de l'art. rencontre avec des sociétés d'auteurs ». Jeanne Revel, Vacarme, 2001/2 (n° 15), p. 31  
<https://www.cairn.info/revue-vacarme-2001-2-page-31.htm>
- 110 Avec ChatGPT et l'irruption de l'intelligence artificielle, la question de la raréfaction du travail et de l'avenir des retraites est relancée. Jean-Michel Bezat, Le Monde, 30 janvier 2023  
[https://www.lemonde.fr/idees/article/2023/01/30/avec-chatgpt-et-l-irruption-de-l-intelligence-artificielle-la-question-de-la-rarefaction-du-travail-et-de-l-avenir-des-retraites-est-relancee\\_6159853\\_3232.html](https://www.lemonde.fr/idees/article/2023/01/30/avec-chatgpt-et-l-irruption-de-l-intelligence-artificielle-la-question-de-la-rarefaction-du-travail-et-de-l-avenir-des-retraites-est-relancee_6159853_3232.html)
- 111 « Economiquement, l'enjeu est très concret puisqu'il pose la question de l'éventuelle rémunération du travail publié sur Internet, alors que ces dispositifs qui les réutilisent font l'objet d'une industrialisation croissante... Plusieurs actions en justice ont déjà été intentées à l'encontre de concepteurs de systèmes d'IA ». Quelles régulations pour la conception des IA génératives ? Erevan Malroux, CNIL, 26 avril 2023  
<https://linc.cnil.fr/fr/dossier-ia-generative-quelles-regulations-pour-la-conception-des-ia-generatives>
- Intelligence artificielle : demain, tous chômeurs ? Gabriel Grésillon, les Echos, 24 Mai 2023  
<https://www.lesechos.fr/tech-medias/intelligence-artificielle/intelligence-artificielle-demain-tous-chomeurs-1945929>
- IA, inflation : 14 millions d'emplois dans le monde vont se volatiliser d'ici 2027, selon un nouveau rapport. Neïla Beyler, Les Echos, 1<sup>er</sup> Mai 2023  
<https://www.lesechos.fr/monde/enjeux-internationaux/14-millions-demplois-dans-le-monde-vont-se-volatiliser-dici-2027-selon-un-nouveau-rapport-1939583>
- Generative AI could raise global GDP by 7%. Goldman Sachs, 5 avril 2023  
<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>
- 112 Pascal Savoldelli. Plateformisation du travail : agir contre la dépendance économique et sociale. Rapport Sénat, 29 septembre 2021  
<https://www.senat.fr/rap/r20-867/r20-8673.html>

113 Comment l'IA générative transforme l'espace NFT. Okx, 25 avril 2023

<https://www.okx.com/learn/generative-ai-nfts>

114 Intelligence artificielle : les photographes réclament des garde-fous. Nicole Vulser, Le Monde, 2 juillet 2023

[https://www.lemonde.fr/economie/article/2023/07/02/intelligence-artificielle-les-photographes-reclament-des-garde-fous\\_6180227\\_3234.html](https://www.lemonde.fr/economie/article/2023/07/02/intelligence-artificielle-les-photographes-reclament-des-garde-fous_6180227_3234.html)

115 « On parle de nouveaux profils, de nouvelles embauches, mais ça représentera un rapport de 1 à 5 entre la personne engagée et tous ceux qui devront partir. » La publicité plonge dans l'intelligence artificielle. Véronique Richebois et al., Le Monde, 11 juin 2023

[https://www.lemonde.fr/economie/article/2023/06/11/la-publicite-plonge-dans-l-intelligence-artificielle\\_6177183\\_3234.html](https://www.lemonde.fr/economie/article/2023/06/11/la-publicite-plonge-dans-l-intelligence-artificielle_6177183_3234.html)

Les dessous de la campagne Undiz générée par des IA. Isadora Lorient, CB News, 26 avril 2023

<https://www.cbnews.fr/marques/image-dessous-campagne-undiz-generee-ia-76209>

An overview of catastrophic AI risks. Center for AI Safety

<https://www.safe.ai/ai-risk>

Le Forum économique mondial chiffre à 14 millions les pertes nettes d'emplois dans le monde d'ici à 2027. Juline Bouissou, le Monde, 2 mai 2023

[https://www.lemonde.fr/economie/article/2023/05/02/le-forum-economique-mondial-chiffre-a-14-millions-les-pertes-nettes-d-emplois-dans-le-monde-d-ici-a-2027\\_6171806\\_3234.html](https://www.lemonde.fr/economie/article/2023/05/02/le-forum-economique-mondial-chiffre-a-14-millions-les-pertes-nettes-d-emplois-dans-le-monde-d-ici-a-2027_6171806_3234.html)

De ChatGPT à Midjourney, les intelligences artificielles génératives s'installent dans les entreprises. Jules Thomas, Alexandre Piquard, Nicolas Six, Corentin Lamy, Louis Adam, Vincent Fagot et Julia Paret, le Monde, 25 avril 2023

[https://www.lemonde.fr/economie/article/2023/04/25/de-chatgpt-a-midjourney-les-intelligences-artificielles-generatives-s-installent-dans-les-entreprises\\_6170873\\_3234.html](https://www.lemonde.fr/economie/article/2023/04/25/de-chatgpt-a-midjourney-les-intelligences-artificielles-generatives-s-installent-dans-les-entreprises_6170873_3234.html)

« En France, les secrétaires sont environ 800 000. Alors que certains pronostiquaient leur disparition avec la généralisation de la bureautique, leur nombre s'est accru dans la seconde moitié des années quatre-vingt pour ensuite ne régresser que modérément au cours des années quatre-vingt-dix » Secrétaire, un métier qui s'est adapté à la généralisation de la bureautique. Ministère de l'Emploi et de la Solidarité, Premières Synthèses, n°8, février 2000.

[https://dares.travail-emploi.gouv.fr/sites/default/files/08ada26b40bb136f65b35a0c35df40/Dares\\_Secr%C3%A9taire%20un%20m%C3%A9tier%20qui%20s%27est%20adapt%C3%A9%20%C3%A0%20la%20g%C3%A9n%C3%A9ralisation%20de%20la%20bureautique.pdf](https://dares.travail-emploi.gouv.fr/sites/default/files/08ada26b40bb136f65b35a0c35df40/Dares_Secr%C3%A9taire%20un%20m%C3%A9tier%20qui%20s%27est%20adapt%C3%A9%20%C3%A0%20la%20g%C3%A9n%C3%A9ralisation%20de%20la%20bureautique.pdf)

116 Discours de Björn Ulvaeus à l'Assemblée Générale de la CISAC 2023, 2 juin 2023

<https://www.cisac.org/fr/Actus-Media/articles/discours-de-bjorn-ulvaeus-lassemblee-generale-de-la-cisac-2023>

117 Complexité. Critique d'une idéologie contemporaine. Sophie Chassat, Fondapol, Juin 2023

118 Bisbille au sommet entre Mark Zuckerberg et Elon Musk, autour de l'IA. Perrine Signoret, L'express, 25 juillet 2017

[https://www.lexpress.fr/economie/high-tech/bisbille-au-sommet-entre-mark-zuckerberg-et-elon-musk-autour-de-l-ia\\_1930301.html](https://www.lexpress.fr/economie/high-tech/bisbille-au-sommet-entre-mark-zuckerberg-et-elon-musk-autour-de-l-ia_1930301.html)

119 Elon Musk calls Bill Gates' understanding of AI 'limited'. Grace Kay, Business Insider, 27 mars 2023

<https://www.businessinsider.com/elon-musk-slams-bill-gates-understanding-ai-chatgpt-openai-2023-3?r=US&IR=T>

120 Complexité. Critique d'une idéologie contemporaine. Sophie Chassat, Fondapol, Juin 2023

121 University of London Press v. University Tutorial Press. Court of Chancery Division, United Kingdom, 1916

<https://www.theipmatters.com/post/university-of-london-press-v-university-tutorial-press>

122 US Copyright Office denies protection for another AI-created image. Blake Brittain, Reuters, 7 septembre 2023

<https://www.reuters.com/legal/litigation/us-copyright-office-denies-protection-another-ai-created-image-2023-09-06>  
<https://fingfx.thomsonreuters.com/gfx/legaldocs/byprqkxpe/AI%20COPYRIGHT%20REGISTRATION%20decision.pdf>

123 Résolution du Parlement européen du 20 octobre 2020 sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle (2020/2015(INI))

[https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277\\_FR.html](https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_FR.html)

- 124 MEMORANDUM OF AGREEMENT FOR THE 2023 WGA THEATRICAL AND TELEVISION BASIC AGREEMENT, 25 septembre 2023, p.69.  
<https://www.wgacontract2023.org/WGAContract/files/Memorandum-of-Agreement-for-the-2023-WGA-Theatrical-and-Television-Basic-Agreement.pdf>
- 125 Les créateurs et artistes interprètes du monde entier réclament le respect de leurs droits face à la prolifération de l'IA. Cisac, 20 juillet 2023  
<https://www.cisac.org/fr/Actus-Media/articles/les-createurs-et-artistes-interpretes-du-monde-entier-reclament-le-respect-de>
- 126 AI learned from their work. Now they want compensation. Gerrit de Vynck, Washington Post, 16 juillet 2023  
<https://www.washingtonpost.com/technology/2023/07/16/ai-programs-training-lawsuits-fair-use>
- 127 'Plagiarism machines': Hollywood writers and studios battle over the future of AI. Dawn Chmielewski and Lisa Richwine, Reuters, 3 mai 2023  
<https://www.reuters.com/technology/plagiarism-machines-hollywood-writers-studios-battle-over-future-ai-2023-05-03>
- 128 Les créateurs et artistes interprètes du monde entier réclament le respect de leurs droits face à la prolifération de l'IA. Cisac, 20 Juillet 2023  
<https://www.cisac.org/fr/Actus-Media/articles/les-createurs-et-artistes-interpretes-du-monde-entier-reclament-le-respect-de>
- 129 Comprendre l'intelligence artificielle. Michel Volle, 30 avril 2017  
<http://michelvolle.blogspot.com/2017/04/comprendre-lintelligence-artificielle.html>
- 130 <https://www.lopinion.fr/secret-defense/quand-de-gaulle-defendait-la-langue-francaise-dans-les-armees>
- 131 « *L'ancienne fortune domaniale est une forme caduque de la richesse, qui a cessé d'avoir sa raison d'être. Elle était la stagnation même de l'argent, dont nous avons décuplé la valeur, en le jetant dans la circulation, et par le papier-monnaie, et par les titres de toutes sortes, commerciaux et financiers. C'est ainsi que le monde va être renouvelé, car rien n'était possible sans l'argent, l'argent liquide qui coule, qui pénètre partout* ».  
Emile Zola. L'Argent  
[https://fr.wikisource.org/wiki/Page:Emile\\_Zola\\_-\\_L%E2%80%99Argent.djvu/132](https://fr.wikisource.org/wiki/Page:Emile_Zola_-_L%E2%80%99Argent.djvu/132)
- 132 Libérer le potentiel des industries culturelles et créatives. Commission Européenne, 27 avril 2010  
<https://op.europa.eu/fr/publication-detail/-/publication/1cb6f484-074b-4913-87b3-344ccf020eef/language-fr>
- 133 Fabio Benedetti Vantino. Droits voisins, les engagements de Google deviennent obligatoires. Les Echos, 21 Juin 2022  
<https://www.lesechos.fr/tech-medias/medias/droits-voisins-google-evite-une-nouvelle-amende-en-soldant-son-contentieux-avec-les-editeurs-de-presse-1414744>
- 134 Google Search  
<https://www.google.com/intl/fr/search/howsearchworks/how-search-works/ranking-results>
- 135 Adam Mosseri. Bringing people closer together. Company News, Jan 11, 2018  
<https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together>
- 136 Nozha Boujema. Les données de confiance, talon d'achille de l'intelligence artificielle. Le Monde, 18 Mai 2022.  
[https://www.lemonde.fr/sciences/article/2022/05/18/les-donnees-de-confiance-talon-d-achille-de-l-intelligence-artificielle\\_6126638\\_1650684.html](https://www.lemonde.fr/sciences/article/2022/05/18/les-donnees-de-confiance-talon-d-achille-de-l-intelligence-artificielle_6126638_1650684.html)
- 137 Unpaid bloggers' lawsuit versus Huffington Post tossed. Jonathan Stempel, reuters, 30 mars 2012  
<https://www.reuters.com/article/us-aol-huffingtonpost-bloggers-idUSBRE82T17L20120330>
- 138 « *Les utilisateurs, bénéficiaires d'un service rendu, deviennent ainsi des quasi collaborateurs, bénévoles, des entreprises. Collecter, stocker et traiter pour être intégré en temps réel à la chaîne de production, les données issues de leur « travail gratuit » contribue à brouiller la frontière entre production et consommation. Attiré par la qualité des interfaces et les effets de réseau, les utilisateurs de Vienne, à travers ses données, des auxiliaires de la production et créer une valeur générant des bénéfices sur les différentes phases de modèle d'affaires. Il est possible de mettre en place une fiscalité incitative fondée sur l'exploitation par les entreprises des données qu'elles collectent via un suivi régulier et systématique de l'activité des utilisateurs de leurs applications* ». Rapport Colin et Collin, 2013.  
[https://www.economie.gouv.fr/files/rapport-fiscalite-du-numerique\\_2013.pdf](https://www.economie.gouv.fr/files/rapport-fiscalite-du-numerique_2013.pdf)

La taxe GAFA instaurée en 2019 rapporte 700 M€ Les recettes de la taxe Gafa ont plus que doublé depuis sa création. Isabelle Couet, Les Echos, 25 octobre 2022

<https://www.lesechos.fr/economie-france/budget-fiscalite/les-recettes-de-la-taxe-gafa-ont-plus-que-double-depuis-sa-creation-1872832>

Taxe sur certains services fournis par les grandes entreprises du secteur numérique (Articles 299 à 300). Code Général des Impôts

[https://www.legifrance.gouv.fr/codes/article\\_lc/LEGIARTI000046868449](https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000046868449)

L'Europe adopte enfin l'impôt minimum à 15 % pour les entreprises. Richard Hiault, Les Echos, 16 décembre 2022

<https://www.lesechos.fr/monde/europe/leurope-adopte-enfin-limpot-minimum-a-15-pour-les-entreprises-1889964>

« La réforme de la taxation des profits générés par les activités numériques des multinationales est dans l'impasse. 'Reconnaissons-le, aujourd'hui, les choses sont bloquées, notamment par les Etats-Unis, l'Arabie saoudite et l'Inde', a déploré lundi le ministre des Finances français, Bruno Le Maire » Echec en vue pour la taxe numérique de l'OCDE. Richard Hiault, Les Echos, 20 février 2023

<https://www.lesechos.fr/monde/enjeux-internationaux/echec-en-vue-pour-la-taxe-numerique-de-locde-1908304>

139 Jean Tirole. Economie du bien commun. PUF 2016, p.534

140 Le « Web3 » expliqué en quatre questions. Olivier Clairouin, Le Monde, 20 avril 2022

[https://www.lemonde.fr/pixels/article/2022/04/20/le-Web3-explique-en-quatre-questions\\_6122944\\_4408996.html](https://www.lemonde.fr/pixels/article/2022/04/20/le-Web3-explique-en-quatre-questions_6122944_4408996.html)

141 #FairlyShare contre le péché originel d'Internet. Vincent Lorphelin, Gilles Babin et al, JDN, 2 septembre 2014

<https://www.journaldunet.com/ebusiness/e-net/1142729-fairlyshare-contre-le-peche-origine-d-internet>

142 w-HA a généré plus de 170 millions d'euros en micro-paiement. Les Echos, 22 avril 2006

<https://www.lesechos.fr/2006/02/w-ha-a-genere-plus-de-170-millions-deuros-en-micro-paiement-563073>

143 Acheter sa musique en ligne sans fausses notes. Le Figaro, 12 octobre 2005

[https://www.lefigaro.fr/secteur/high-tech/2005/10/12/32001-20051012ARTFIG90020-acheter\\_sa\\_musique\\_en\\_ligne\\_sans\\_fausses\\_notes\\_.php](https://www.lefigaro.fr/secteur/high-tech/2005/10/12/32001-20051012ARTFIG90020-acheter_sa_musique_en_ligne_sans_fausses_notes_.php)

144 Twitter to let publishers charge users per article read, says Elon Musk. The Guardian, 30 avril 2023

<https://www.theguardian.com/technology/2023/apr/30/twitter-to-let-publishers-charge-users-per-article-read-says-elon-musk>

145 Bitcoin is still the future of payments, says Lightspark CEO David Marcus. Nilay Patel, The Verge, 25 avril 2023

<https://www.theverge.com/23696466/bitcoin-crypto-lightning-network-lightspark-ceo-david-marcus>

Cinq choses à savoir sur Lightspark, la start-up de David Marcus. Charlie Perreau. Les Echos, 3 juillet 2023

<https://www.lesechos.fr/start-up/deals/cinq-choses-a-savoir-sur-lightspark-la-start-up-de-david-marcus-1958235>

146 CISAC, rapport annuel 2023

<https://www.cisac.org/fr/Actus-Media/news-releases/la-cisac-publie-son-rapport-annuel-2023>

CISAC and JASRAC discuss initiatives to bring creators fair remuneration in the digital world. Cisac, 28 avril 2023

<https://www.cisac.org/Newsroom/news-releases/cisac-and-jasrac-discuss-initiatives-bring-creators-fair-remuneration>

147 Thomas Pontiroli. Après l'échec du métavers, son créateur veut reprendre les choses en main. Les Echos, 11 juillet 2023

<https://www.lesechos.fr/tech-medias/hightech/apres-lechec-du-metavers-son-createur-veut-reprendre-les-choses-en-main-1961023>

148 Art Blocks Debuts NFT Marketplace With Enforced Creator Royalties, Andrew Hayward, Decrypt, 29 mars 2023

<https://decrypt.co/124848/art-blocks-nft-marketplace-enforced-creator-royalties>

Solana imagine démocratiser la blockchain avec ChatGPT. Thomas Pontiroli, 25 mai 2023

<https://www.lesechos.fr/tech-medias/intelligence-artificielle/solana-imagine-democratiser-la-blockchain-avec-chatgpt-1946555>

L'économie mondiale est sur le point d'être transformée avec la cryptomonnaie de Facebook. Vincent Lorphelin, Le Monde, 19 juin 2019

[https://www.lemonde.fr/idees/article/2019/06/19/l-economie-mondiale-est-sur-le-point-d-etre-transformee-avec-la-cryptomonnaie-de-facebook\\_5478275\\_3232.html](https://www.lemonde.fr/idees/article/2019/06/19/l-economie-mondiale-est-sur-le-point-d-etre-transformee-avec-la-cryptomonnaie-de-facebook_5478275_3232.html)

Avec la cryptomonnaie libra, Facebook aspire à un rôle central dans le nouveau théâtre de la transition numérique. Vincent Lorphelin, Le Monde, 1<sup>er</sup> juillet 2020

[https://www.lemonde.fr/idees/article/2020/07/01/avec-la-libra-facebook-entend-prendre-un-role-central-dans-le-nouveau-theatre-de-la-transition-numerique\\_6044789\\_3232.html](https://www.lemonde.fr/idees/article/2020/07/01/avec-la-libra-facebook-entend-prendre-un-role-central-dans-le-nouveau-theatre-de-la-transition-numerique_6044789_3232.html)

149 Pierre-Olivier Beffy, Jean-Marc Béguin, Pierre-Jean Benghozi, Laurent Bloch, Hugues Chevalier, Vincent Lorphelin et Michel Volle, « L'économie : un modèle de l'économie numérique », Revue d'économie industrielle [En ligne], 165 | 1<sup>er</sup> trimestre 2019, mis en ligne le 1<sup>er</sup> janvier 2022

<http://journals.openedition.org/rei/7998>

Le Triomphe de l'économie de l'utilité. Vincent Lorphelin, Christian Saint-Etienne et Michel Volle, Le Monde, 18 juin 2015

[https://www.lemonde.fr/idees/article/2015/06/19/le-triomphe-de-l-economie-de-l-utilite\\_4657975\\_3232.html](https://www.lemonde.fr/idees/article/2015/06/19/le-triomphe-de-l-economie-de-l-utilite_4657975_3232.html)

150 ChatGPT, nouvel horizon pour les scénaristes. Djaïd Lamak, Le Monde, 25 avril 2023

[https://www.lemonde.fr/culture/article/2023/04/25/chatgpt-nouvel-horizon-pour-les-scenaristes\\_6170983\\_3246.html](https://www.lemonde.fr/culture/article/2023/04/25/chatgpt-nouvel-horizon-pour-les-scenaristes_6170983_3246.html)

151 « N'oublions pas l'intelligence collective ». Vincent Lorphelin, Le Monde, 2 janvier 2018

[https://www.lemonde.fr/pixels/article/2018/01/02/n-oublions-pas-l-intelligence-collective\\_5236747\\_4408996.html](https://www.lemonde.fr/pixels/article/2018/01/02/n-oublions-pas-l-intelligence-collective_5236747_4408996.html)

Intelligence as a planetary scale process. Adam Frank, David Grinspoon Sara Walker, Cambridge University Press, 7 février 2022

<https://www.cambridge.org/core/journals/international-journal-of-astrobiology/article/intelligence-as-a-planetary-scale-process/5077C-784D7FAC55F96072F7A7772C5E5>

152 Quantin, Jean-Louis. Les institutions de censure religieuse en France (XVI<sup>e</sup>- XVII<sup>e</sup> siècles) In : Hétérodoxies croisées. Catholicismes pluriels entre France et Italie, XVI<sup>e</sup>-XVII<sup>e</sup> siècles [en ligne]. Rome : Publications de l'École française de Rome, 2015 (généré le 12 novembre 2023). ISBN : 9782728311446.

<https://books.openedition.org/efr/2837>

153 Louis Figuier, « L'hygiène et les chemins de fer », L'Année scientifique et industrielle, Paris, Hachette, 1863, p. 389-396

154 Vincent Lorphelin, Le Rebond Economique de la France, Pearson Village Mondial 2012

155 Baudelaire - Curiosités esthétiques, p.260

[https://fr.wikisource.org/wiki/Page:Baudelaire\\_-\\_Curiosit%C3%A9s\\_esth%C3%A9tiques\\_1868.djvu/267](https://fr.wikisource.org/wiki/Page:Baudelaire_-_Curiosit%C3%A9s_esth%C3%A9tiques_1868.djvu/267)

156 Citizen's Basic Income: The Answer is Blowing in Wind. Eduardo Matarazzo Suplicy, 2006

<https://usbig.net/papers/152-Suplicy-blowin.pdf>

157 Réforme audiovisuelle : les plates-formes comme Netflix devront investir 25 % de leur chiffre d'affaires en France. Marina Alcaraz, Les Echos, 14 janvier 2020

<https://www.lesechos.fr/tech-medias/medias/reforme-audiovisuelle-les-plates-formes-comme-netflix-devront-investir-25-de-leur-chiffre-daffaires-en-france-1162805>

158 Directive (UE) 2018/1808 du Parlement Européen et du Conseil, 18 novembre 2018

<https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32018L1808>

Les producteurs audiovisuels qui travaillent pour Netflix ou Disney Channel auront accès aux aides du CNC. Nicole Vulser, Le Monde, 15 décembre 2022

[https://www.lemonde.fr/economie/article/2022/12/15/les-producteurs-audiovisuels-qui-travaillent-pour-netflix-ou-disney-channel-auront-acces-aux-aides-du-cnc\\_6154604\\_3234.html](https://www.lemonde.fr/economie/article/2022/12/15/les-producteurs-audiovisuels-qui-travaillent-pour-netflix-ou-disney-channel-auront-acces-aux-aides-du-cnc_6154604_3234.html)

