La solution Controv3rse

Rémunérer les ayant-droits par l'IA générative pour valoriser le capital humain et culturel de l'Europe

CONTROV3RSE

Mission du CSPLA. 9 décembre 2024. Rapport intermédiaire.

Contexte

1 – Objet

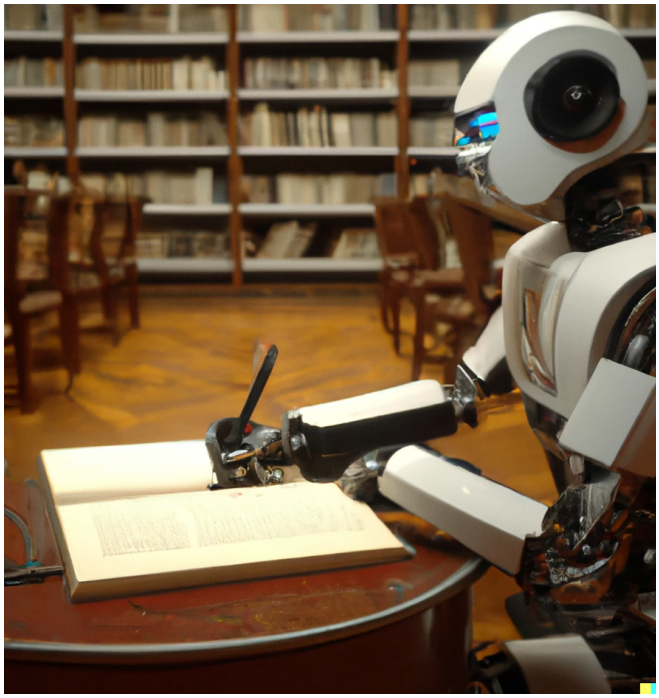
Dans le conflit qui oppose auteurs et exploitants d'IA générative, les premiers crient au pillage[1], les seconds opposent leur bon droit. Le « fair use » permet en effet à tous de feuilleter des livres sur Google Books, gratuitement, sans payer les auteurs. Comme les deux positions sont juridiquement défendables, il faudra au moins dix ans pour que les multiples procès en cours aboutissent. Dans l'intervalle, les fournisseurs d'IA générative auront le temps de développer les Uber des données. La filière va se segmenter entre créateurs de données, plateformes, agrégateurs, négociateurs, entraîneurs, affineurs, exploitants et intégrateurs d'IA générative, ce qui va diluer l'acte de création. Pendant dix ans, le patrimoine culturel que constitue aujourd'hui les droits d'auteurs aura été mélangé, diffusé et recyclé dans des contenus dont les traces originelles auront été effacées. Le pouvoir techno-politique, symbolisé par l'alliance Trump-Musk, aura distillé l'idée que les machines ont le « droit d'apprendre » gratuitement. Même si les auteurs finissent par obtenir satisfaction sur le plan juridique, leur compensation se fera au regard des comparables fournis par les Uber des données, c'est-à-dire peu de chose. On se rendra compte rétrospectivement que leur système de défense actuel, fondé sur le seul droit existant, n'aura été qu'une ligne Maginot facile à contourner.

La crainte de ce scénario tendanciel se répand. Les parlementaires européens ont déjà voté l'AI Act, qui réaffirme le principe du droit d'auteur, mais ne répond pas à toutes les questions : en particulier**,** comment le « résumé » des sources évoqué par l'AI act devrait-il être conçu pour aboutir à la rémunération équitable des auteurs ?

Cette note intermédiaire ne prétend pas synthétiser les nombreuses propositions publiées, mais cherche à élucider l'intérêt général par l'enjeu plus global de la valorisation du capital humain et culturel de l'Europe, et proposer une solution « bêta » de mise en œuvre, pour contribuer à éclairer le débat public.

Cette contribution aux travaux du CSPLA se base sur l'état de l'art et de la recherche, pour concilier les contraintes :

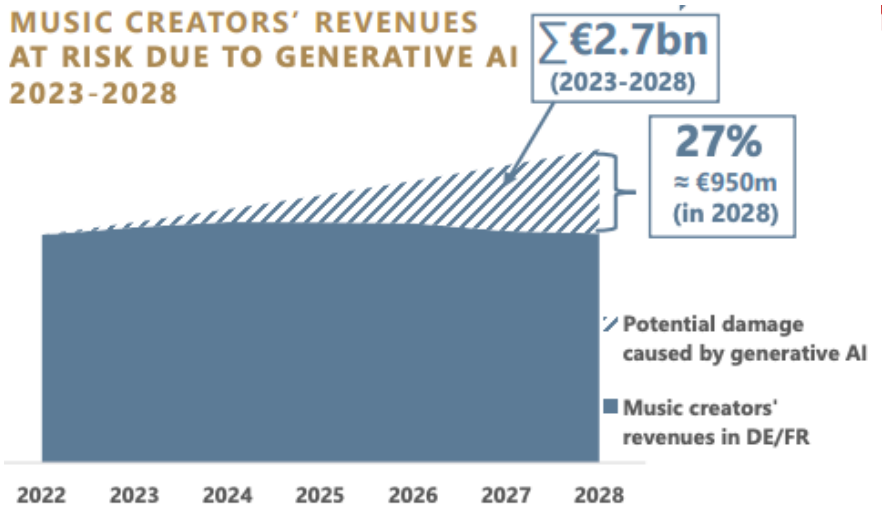
- Juridique : AI act tel que publié[2]
- Juridico-économique : hypothèse de la mise en œuvre d'un droit à rémunération. « La loi décide de compenser un usage parce qu'il crée un préjudice économique. Il y a un prélèvement, ensuite réparti entre les titulaires de droits. Une commission pourrait fixer son montant[3] »
- Sociétale : degré d'équité acceptable pour les auteurs
- Économique : coût de mise en œuvre acceptable pour les fournisseurs d'IA générative et les Organismes de Gestion Collective
- Politique : méthode permettant de surmonter les oppositions et de dessiner une vision pour l'intérêt général

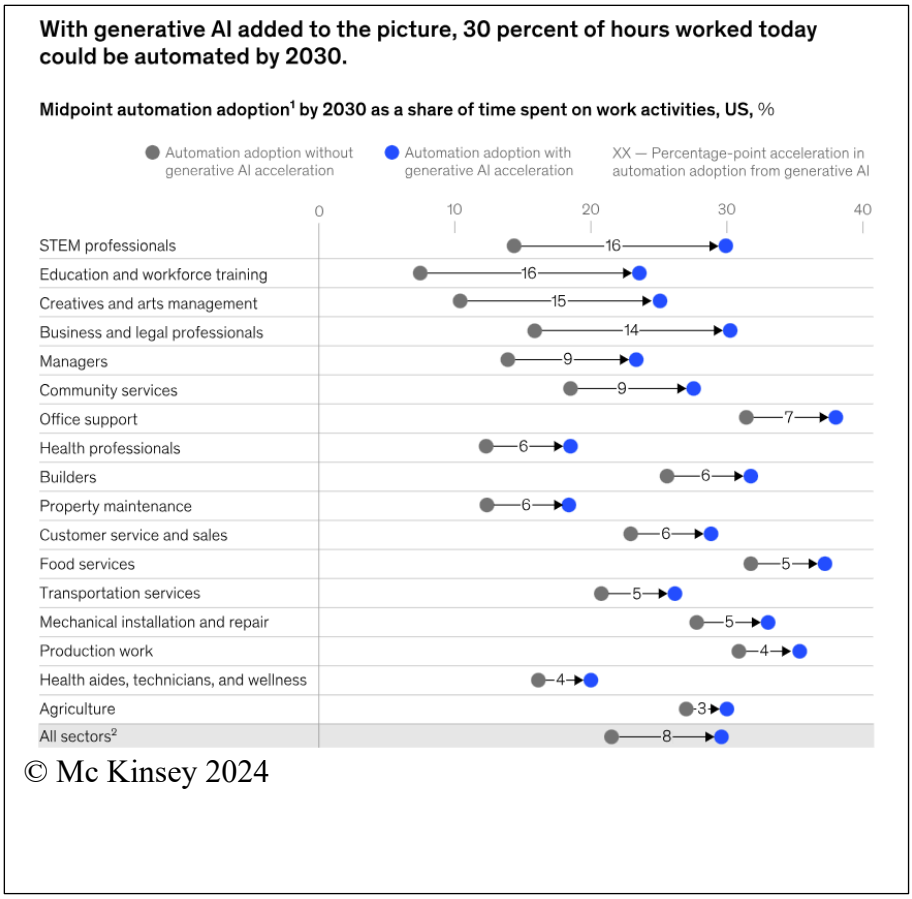


AI-generated by DALL-E

2 – L'IA ubérisante menace le destin européen

En 2028, les créateurs de musique pourraient perdre 27% de leurs revenus à cause de l'IA générative (source Goldmedia, Sacem, Gema)





© Mc Kinsey 2024

Dans le conflit de l'IA générative, l'intérêt économique des Etats-Unis pour la révolution de l'IA pèsera au bénéfice des exploitants[4]. L'Europe, qui a identifié depuis la stratégie de Lisbonne l'étendue de son patrimoine culturel, voudra symétriquement le protéger.

L'enjeu de cette bataille dépasse donc l'antagonisme des auteurs et des exploitants. Il concerne l'économie et la culture européenne, son influence, son poids dans le monde, c'est-à-dire son destin. La pépite Mistral AI est une startup française qui, 18 mois seulement après sa naissance, pèse 6 milliards d'euros. Cet exploit pourrait d'ores et déjà la placer dans le CAC40. Pourtant, c'est une poussière face aux Gafam, dont la capitalisation individuelle dépasse 1000 milliards. Le rapport de Mario Draghi[5] met l'IA au centre des enjeux macro-économiques. Il révèle que l'Europe a 8 fois moins de licornes que les Etats-Unis et un capital-risque 5 fois moindre[6]. Avec l'IA générative, l'écart va encore se creuser. La puissance de l'IA ubérisante des grandes plateformes devient irrésistible, au-delà de la créativité, pour l'artisanat, les métiers d'art et expertises, dans toutes les filières économiques. Les auteurs sont en première ligne, mais tous les métiers vont suivre[7], y compris les chercheurs (entretien avec Michelle Bergadaà ci-dessous).

La recette de l'IA ubérisante est simple. Personnaliser par exemple l'IA générative d'entreprise avec toutes ses archives de données permet d'obtenir un brouillon pour chaque nouveau message, compte-rendu, projet, présentation. Les rectifications des professionnels entraînent l'IA jusqu'à ce que le taux d'erreur devienne acceptable. Les gains de productivité sont sensibles mais, en contrepartie, les positions de l'entreprise sur les chaînons stratégiques de la création de valeur diminuent au bénéfice de l'exploitant de l'IA.

Comme tous ces savoir-faire sont le croisement de savoirs, de pratiques, de matériaux d'organisation sociale, c'est le tissu des entreprises qui est en jeu[8]. C'est comme si, au moment de la révolution industrielle, les américains avaient été les seuls propriétaires au monde des machines à vapeur et contrôlaient l'organisation tayloriste du travail. En accordant des capitalisations en milliers de milliards, les bourses ne valorisent plus aujourd'hui des entreprises, mais des monopoles mondiaux.

L'Europe dispose encore d'un capital humain[9] et culturel considérables. Miroir de l'humanité et de l'esthétique, fondement de nos identités et de nos communautés, c'est une source infinie d'inspiration et d'innovation pour l'économie et le développement durable de l'UE[10]. Ce bien commun est menacé d'extraction par l'IA ubérisante. L'Europe semble condamnée à une « lente agonie », dixit Mario Draghi, car refuser ces technologies c'est se condamner à un décrochage durable, comme celui de la Chine pendant deux siècles. L'accepter sans contrepartie, c'est se laisser coloniser sans combattre.

Michelle Bergadaà, Professeure émérite à l'université de Genève.
Co-autrice de *Réinventer l'intégrité académique à l'ère de l'intelligence artificielle (à paraître, EMS, février 2025)*

Vous dites que l'AI Act ne concerne pas que les auteurs, mais aussi les chercheurs, pourquoi ?

Nous devons réfléchir à appliquer l'AI Act pour rémunérer les auteurs de « recherches authentiques ». La France paye mal ses chercheurs, cela freinerait leur fuite vers l'étranger ou le privé. Cela coûte bien plus cher à la société d'écraser sous la bureaucratie ses jeunes chercheurs que de bien les payer. Beaucoup se démotivent et font des heures supplémentaires d'enseignement au lieu de faire de la recherche. Alors si l'Etat ne peut pas les rémunérer convenablement, il faut bien que le système de production des articles, livres, rapports le fasse. Cela doit être la pierre angulaire de la stratégie européenne de l'IA générative. Avec en prime, un système automatisé de détection des fraudes.

La fraude n'a-t-elle pas toujours existé ?

La crise des opioïdes aux USA [Le fentanyl et apparentés, prescrits par ordonnance, ont à eux seuls été responsables d'environ 71.000 décès par overdose de drogue en 2021] est arrivée par un seul et unique article faux, mal cité et repris. C'était bien avant l'IA. Avec l'IA, la diffusion des articles frauduleux se fait essentiellement via de nouveaux articles. C'est un nouveau modèle d'essaimage du faux.

Aujourd'hui trois temporalités cohabitent. Premièrement le temps instantané et émotionnel (de Trump et de la génération Z) permet de produire de manière non seulement des fausses informations, mais aussi de faux auteurs, tout aussi crédibles. Deuxièmement l'éternel présent des communautés de jadis entretient un monde alternatif où la terre est plate et où la javel élimine la Covid. Troisièmement le temps linéaire est poussé par la technologie matérialiste d'Elon Musk. L'IA est un accélérateur de fraude qui foisonne dans les interstices de ces temporalités.

Comment agir ?

Les initiatives comme Pubpeer et Retraction Watch Database se trompent de niveau. Elles cherchent à coincer des délinquants individuels, un par un. Ensuite deux ans sont nécessaires pour boucler l'enquête, par exemple par une commission d'éthique du CNRS, et encore un an pour obtenir la rétractation dans les journaux.

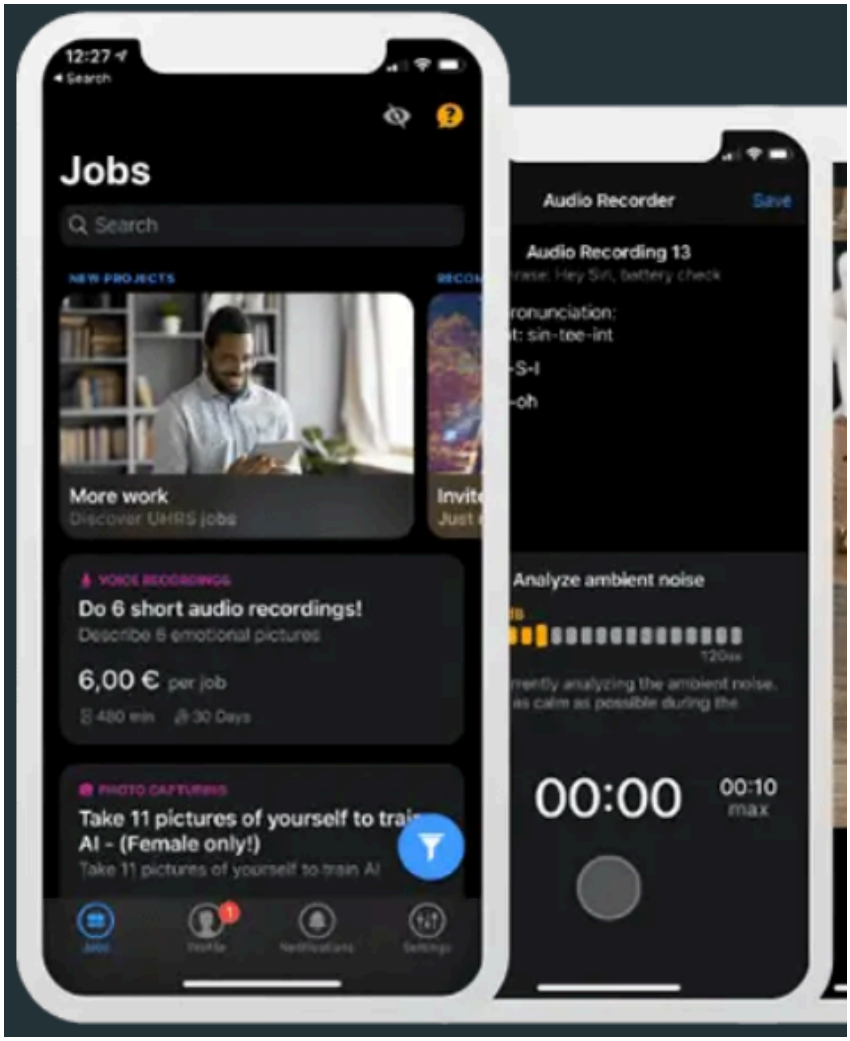
Il faut appréhender ce phénomène de manière systématique avec des radars d'intégrité, une injonction pour les fraudeurs de rendre l'argent perçu, et une exclusion des bases de données de l'IA. Le futur de notre société, c'est les jeunes que nous formons. Aidons-les par notre posture ferme et en rémunérant leurs créations originales par l'IA générative. Cela motiverait bien des jeunes chercheurs intègres à créer de la connaissance.

3 – Sacémisation contre ubérisation

Face à l'IA ubérisante, la propriété intellectuelle reste un levier de défense puissant.

Le prix Nobel Jean Tirole nous enseigne que 1/ la richesse des nations dépend de leur capacité à capter la valeur au niveau de l'innovation[11]. 2/ l'innovation provient de l'initiative individuelle, car le propre du travail créatif est que l'on ne sait pas ce qu'on va trouver[12]. 3/ la propriété intellectuelle est un mal nécessaire pour stimuler cette créativité[13]. 4/ empêcher une entreprise de commercialiser ses produits tant qu'elle n'a pas payé ses redevances est une arme très efficace contre les Gafam[14]. 5/ des pools de propriété intellectuelle avec autorisation de licences individuelles leur imposerait d'acquérir des licences[15], mais à un prix concurrentiel[16].

Les OGC (organismes de gestion collective), comme la SACEM, sont fondés sur le principe de la propriété intellectuelle des auteurs. Ils tracent cette propriété, collectent les redevances, et rémunèrent les ayant-droits selon des grilles de répartition prédéfinies.



©Clickworker 2024

A l'inverse, les plateformes de la gig economy sont fondées sur une logique de commande. Clickworker[17] paye directement les utilisateurs 6€ pour décrire 6 images à haute voix et en poster l'enregistrement. Ou pour se prendre en photo en train de faire des exercices physiques. Le résultat sert à entraîner les IA.

Ce modèle de type « Uber des données » est en pleine croissance grâce au besoin massif de données des exploitants d'IA générative, tandis que la SACEM, leader des OGC, interdit l'utilisation des œuvres par l'IA générative.

AI ubérisante et sacémisation sont deux modèles concurrents. Le premier favorise la demande, l'exécution, la marchandisation immédiate, le néo-taylorisme cognitif, la normalisation du travail. Le second favorise l'offre, l'initiative, la propriété intellectuelle, la créativité, la diversité culturelle. Le premier exploite le capital humain et culturel, le second le cultive.

L'Union européenne voulait « devenir l'économie de la connaissance la plus compétitive et la plus dynamique du monde, capable d'une croissance économique durable accompagnée d'une amélioration quantitative et qualitative de l'emploi et d'une plus grande cohésion sociale ».

Cet objectif, défini trop tôt, reprend tout son sens dans le contexte de l'IA générative. Le bien commun du capital humain et culturel pourrait constituer dans cette nouvelle filière une position stratégique de l'Europe. Il faut à cette fin trouver une solution sacémisante moderne, qui puisse concurrencer efficacement l'IA ubérisante.

4 – Une solution sacémisante pour l'Europe

Entretien avec Jean-Paul Betbèze

Economiste, Membre du Comité scientifique de la Fondation Robert Schuman

Quels devraient être les points clés de la "sacémisation" ?

JPB : « Une solution "sacémisante" pour l'Europe devrait :

- Être fondée sur la propriété intellectuelle des auteurs, comme les OGC. Il faut maintenir et garantir la propriété intellectuelle des auteurs, qui pourrait, sinon, être dissoute ou disparaître dans la révolution actuelle de la communication. La communication se mondialise, c'est la base du changement majeur que nous vivons, avec la multiplication des messages en tous sens. L'apport de chacun doit être reconnu, sauf à assécher le mouvement mondial en cours par une concentration excessive au bénéfice de certains auteurs, les autres ne trouvant aucun avantage à leur travail et à leurs idées, aucune reconnaissance, aucune rémunération. À cause des écarts de disponibilité des capitaux, à l'avantage des Etats-Unis pour une grande part, le danger est une monopolisation des sources apparentes de création.

- Collecter des redevances sur l'exploitation des œuvres par l'IA. La "sacémisation" permet non seulement de tracer, mais aussi d'identifier et estimer les sources, même lorsqu'elles semblent minimes. Ce processus permet de conserver toutes les origines et donc toutes les reconnaissances de créations et d'améliorations : rien ne se perd, tout se garde.

- Reconnaître la créativité et la distinguer du contenu amateur. Dans la prolifération des travaux, il importe de pouvoir mettre en avant ceux qui sont originaux, pour les signaler et les valoriser.

- Mettre en œuvre des mécanismes d'incitation individuelle, pour que le prix de la créativité soit compétitif sur le marché et compatible avec le coût de production. Le risque de la multiplication des travaux est la banalisation, d'autant plus que leur coût de production physique est évidemment faible. Une facturation liée au seul coût de production court le risque d'être désincitative. Le grand avantage de la "sacémisation" est de s'éloigner d'une logique de prix liée au coût de production, pour aller vers une autre logique, celle des marchés, qui est fondée sur l'importance du succès, la répétition du message.

Quel est l'avantage d'un marché dynamisé par la "sacémisation" ?

Dans cette logique de marché, c'est la demande mesurée, d'où découle la demande anticipée, qui fabrique les nouveaux prix, donc les marges. Dans ce contexte les marges ne peuvent que croître, donc les profits

marginaux, donc les valorisations. Cette différence entre les coûts faibles, dépistés par les moteurs de recherche et les valorisations qui auront ainsi une base objective, mises à jour par les outils de type Sacem, est la base de la logique proposée. Elle permettra de mettre en avant les apports de chacun, notamment dans un contexte européen.

Et pour les activités créatives ?

N'oublions pas que la concurrence monopolistique est toujours à l'œuvre : elle pousse à la réduction du nombre de concurrents, jusqu'à ce qu'un seul émerge : c'est le monopole. Là plus qu'ailleurs, le gagnant rafle toute la mise. La "sacémisation" doit s'y opposer en permettant des rémunérations mieux réparties avant qu'un seul monopoleur n'apparaisse. Ceci est évidemment favorable à la recherche, à l'innovation, à la création, puisque l'on sait bien que la situation de monopole se retourne contre le monopoleur lui-même, dans une situation de marché qui s'assèche à terme, car sa dynamique conduit à la baisse de ses profits. La "sacémisation" respecte ainsi les créations et freine la sélection impitoyable et, au fond, destructrice du monopole lui-même.

Dans le marché des idées, la vitesse est mortelle, la mémoire des apports est décisive, en liaison avec leur juste rémunération. C'est aujourd'hui possible et, on le comprend, plus nécessaire que jamais. Il s'agit de sauver la création grâce aux progrès qu'offre la technologie. »

Solution

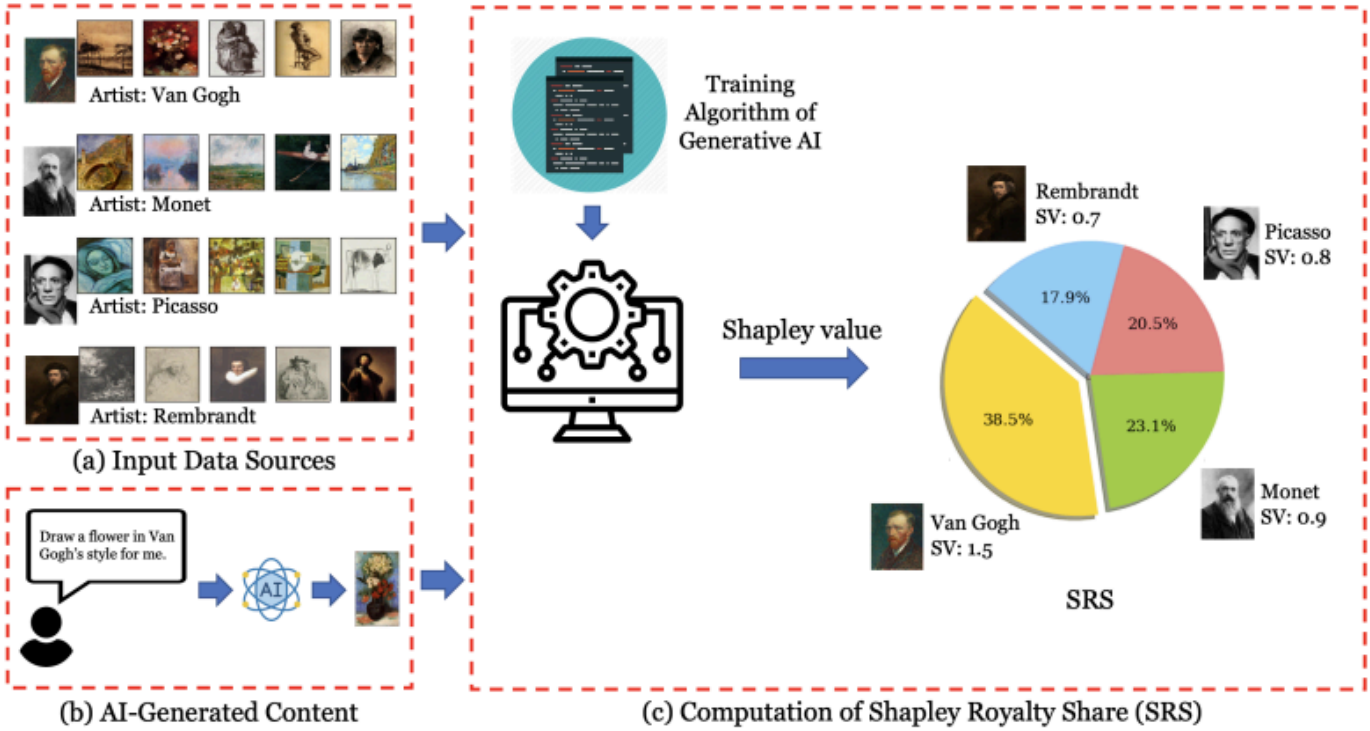
5 - Solution idéale

La solution idéale de la rémunération équitable fait consensus chez les chercheurs. Il s'agit de calculer la valeur de Shapley pour chaque auteur et d'effectuer la répartition des rémunérations au pro rata de cette valeur[18].

En théorie des jeux, Lloyd Shapley a introduit l'idée en 1953 qu'il était possible, dans un jeu coopératif, de calculer une répartition équitable des gains entre les joueurs. On considère par exemple un écosystème marin avec toutes ses espèces vivantes - algues, anémones, coquillages, étoiles de mer – et on mesure l'utilité de chaque espèce par sa contribution à la vie dans l'écosystème. Quand on retire les étoiles de mer, 50% des autres espèces disparaissent, contre 10% quand on retire les anémones. La valeur de Shapley des étoiles de mer est cinq fois supérieure à celle des anémones.

Dans le cas de l'IA générative, la répartition équitable des rémunérations des œuvres d'entraînement devrait idéalement se faire en fonction de l'utilité des œuvres d'entraînement pour les contenus AI-générés, par auteur. Dans la figure 1, la valeur de Shapley de Van Gogh est de 1,5 soit 38,5% de la rémunération totale.

Figure 1 – Valeur de Shapley



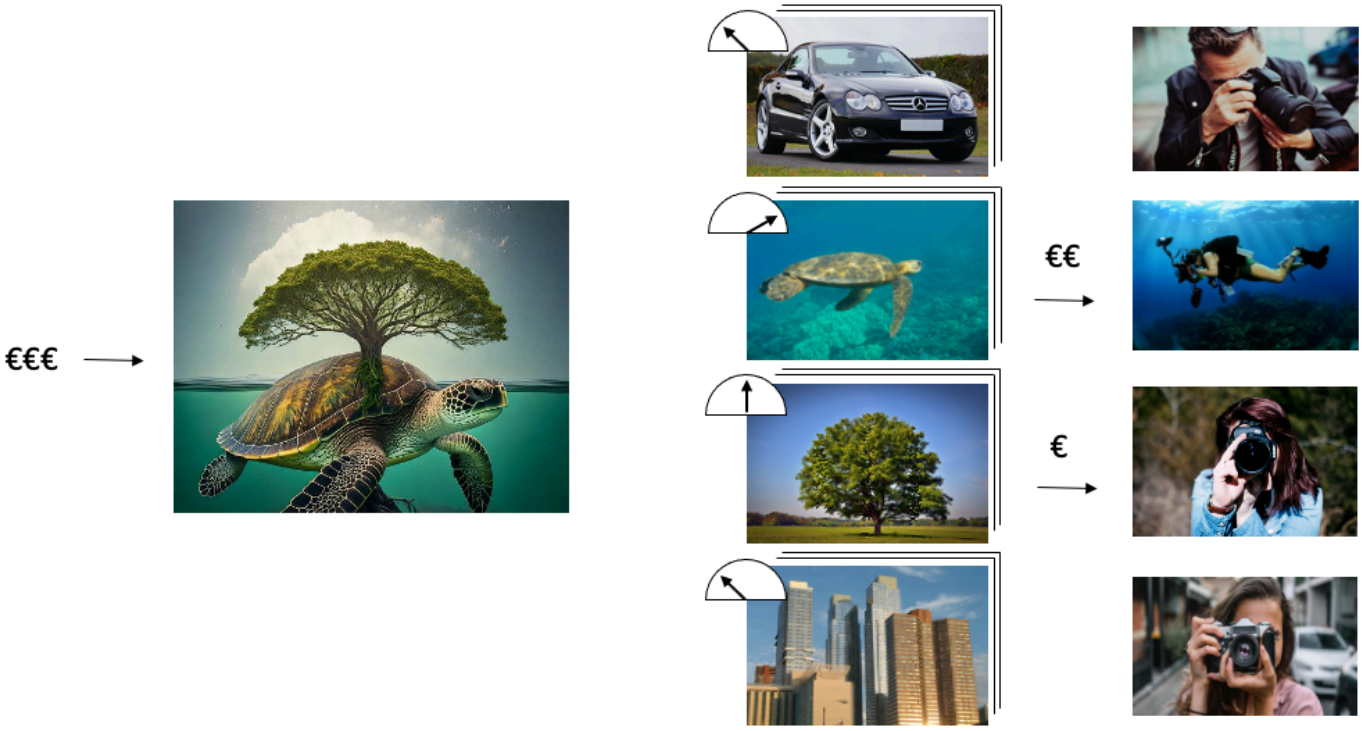
En pratique, le coût de calcul de la valeur de Shapley est prohibitif.

6 – Solution de base

Une autre solution est d'ores et déjà commercialisée pour les contenus d'images.

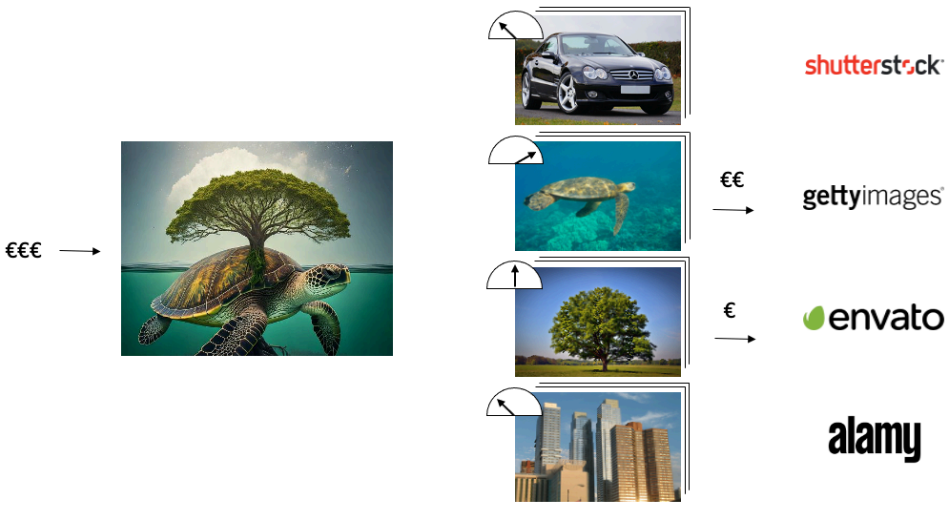
Elle est mise en œuvre par la startup israelo-américaine Bria, en collaboration avec les grandes banques d'images (Shutterstock, Getty, Envato) et intégrée par les Big Tech (Microsoft, Amazon, Nvidia). En première approche, Bria compare le contenu généré avec toutes les images d'entraînement, afin de rémunérer les auteurs de manière proportionnelle à la ressemblance ainsi mesurée (figure 2).

Figure 2 – Measure similarities of generated content and original photos to compensate authors



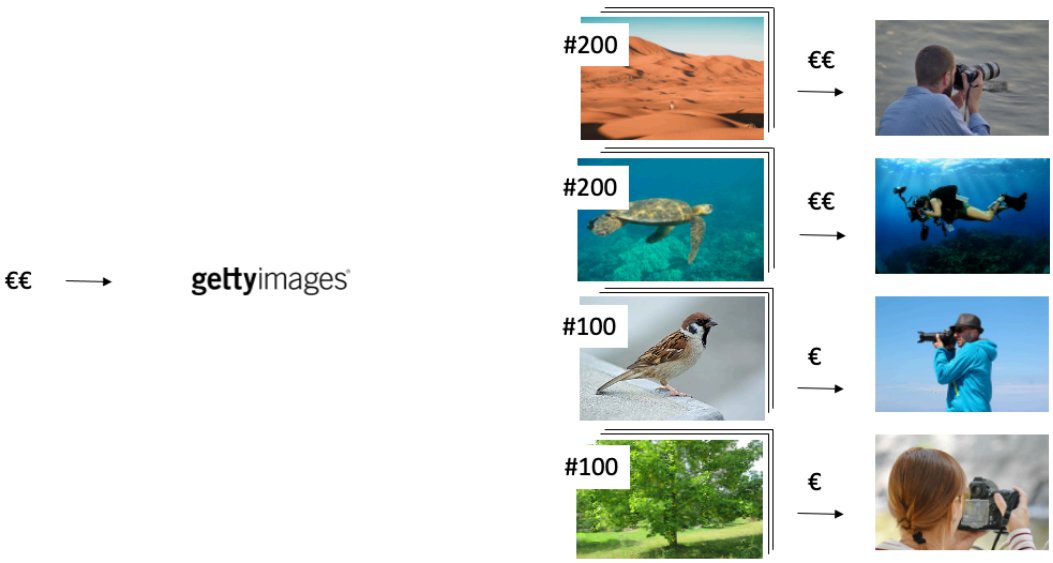
De manière plus précise, il s'agit d'un algorithme qui associe un ensemble d'indicateurs d'utilités (type, style, taille, objets reconnus, mots-clés,…) à chaque image. Une valeur d'utilité de chaque œuvre d'entraînement est calculée pour chaque contenu généré, que les scientifiques nomment « similarité vectorielle dans un espace latent » [19] (voir details plus bas). La somme des scores est calculée par banque d'images (Shutterstock, Getty, Envato, Alamy...), ce qui provoque une première rémunération des banques proportionnelle à ces sommes (figure 3).

Figure 3 – Measure similarities of generated content and original photos to compensate images banks



Ces banques répartissent ensuite cette première rémunération entre les auteurs, en fonction du nombre individuel de leurs œuvres contribuant au dataset vendu (« pay to train »).

Figure 4 – Number of original photos to compensate authors



Quelques affinages sont apportés, les images assorties de métadonnées de qualité reçoivent ainsi un bonus.

En positif, cette solution de base est :

- économiquement compatible avec le modèle d'affaires des IA génératives
- acceptée depuis deux ans par les banques d'images et surtout par les auteurs
- compatible avec une lecture minimaliste de l'obligation des fournisseurs de modèles de « mettre à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner les modèles d'IA à usage général, tout en tenant dûment compte de la nécessité de protéger les secrets d'affaires et les informations commerciales confidentielles »
- en cours d'extension à d'autres types de contenus : 3D, musique, voix, image animée, video[20]

Elle présente cependant plusieurs défauts :

- il n'est pas prouvé que cette solution s'adaptera à tous les types de contenus, par exemple les jeux
- la similarité des contenus original et généré ne mesure pas toujours un lien de causalité
- le degré de contribution d'une image originale est mesuré de manière très approximative. Dans l'exemple de la figure 4, le photographe du désert est rémunéré autant que celui de tortues marines, car leur nombre de contributions au dataset de Getty Images est identique, alors que l'utilité des photos de désert au contenu généré est nul

On notera que :

- même si la répartition de la rémunération pour un contenu généré particulier est approximatif, son équité s'améliore probablement avec la quantité de contenus générés.
- l'algorithme de répartition lui-même n'est pas public. Il est possible que les auteurs réclament dans l'avenir davantage de transparence, mais il est probable aussi, à l'inverse, qu'une dose de secret soit nécessaire pour éviter des biais opportunistes.
- on ne connaît pas non plus le taux moyen de rémunération des auteurs par rapport au chiffre d'affaires des contenus générés qui, à notre avis, pourrait quant à lui être publié.

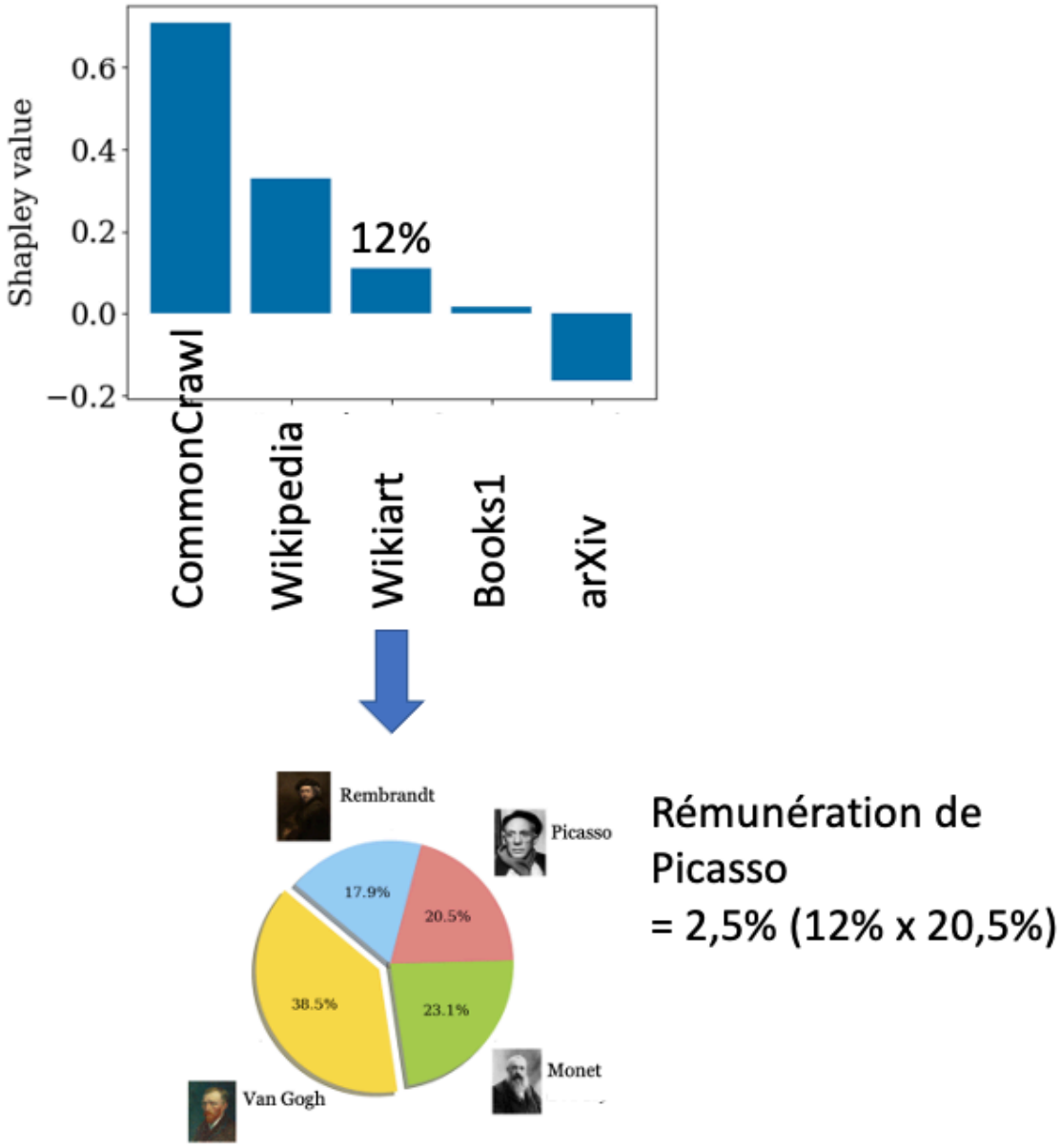
7 - La solution Controv3rse

Nous proposons d'adopter une méthode qui organise l'amélioration constante de la solution de base au regard du principe de l'équité. De manière schématique avant de détailler cette solution :

- la rémunération équitable est collectée auprès des fournisseurs des modèles d'IA génératives (de manière similaire aux discothèques)
- en fonction d'un taux réglementaire (de type Spré – Société pour la Perception de la Rémunération Equitable)
- appliqué à une estimation du chiffre d'affaires des contenus générés à partir du modèle
- la rémunération est d'abord répartie entre les datasets en application d'une première valorisation relative (ex. : similarité vectorielle, Shapley)
- puis par auteur ayant contribué à chaque dataset, en application d'une deuxième valeur relative (ex. : pay-to-train)

La figure 5 illustre la répartition en deux temps. La valorisation relative des datasets permet de répartir la collecte entre ceux-ci, Wikiart recevant 12%. La valeur pay-to-train est ensuite attribuée à chaque auteur ayant contribué au dataset, Picasso recevant une part relative de 20,5%, soit 2,5% = 12% x 20,5% de la collecte totale.

Figure 5 – Solution mixte Shapley/Pay-to-train



Mise en œuvre technique

8 – Valorisation d'un dataset

Similarité de vecteurs

Figure 6 – le degré d'attribution d'une œuvre (image 1) à un contenu généré (image 2) est calculée par un score de similarité entre les deux images, traitées par deux réseaux neuronaux siamois. La similarité est la distance euclidienne entre les deux représentations vectorielles des images[21]

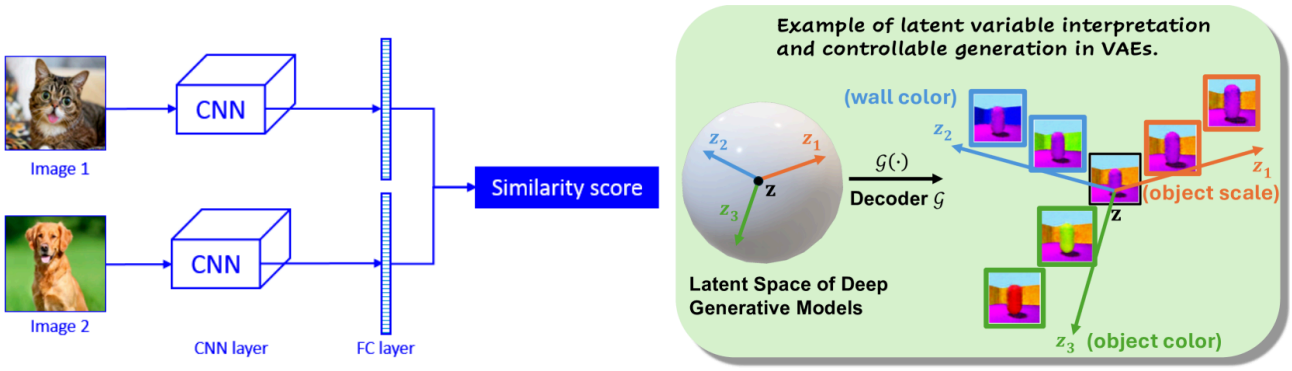
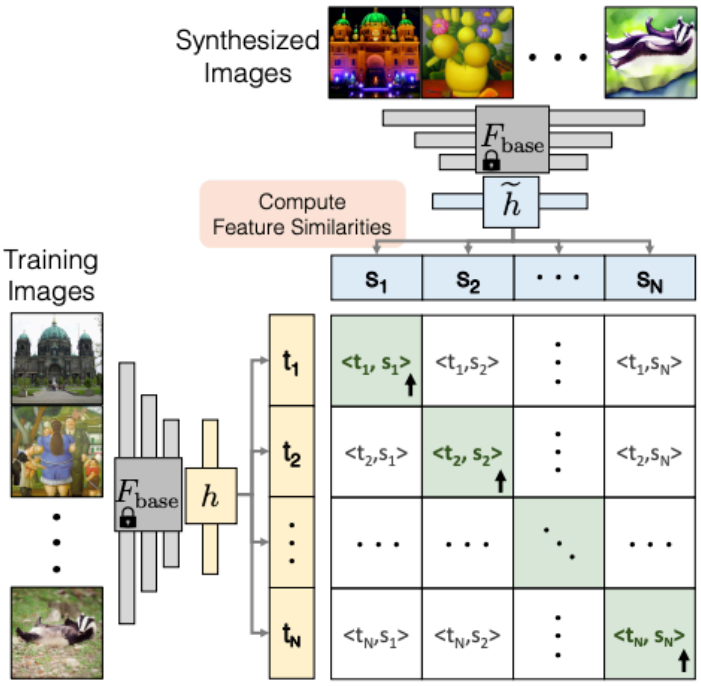


Figure 7 – Training for attribution

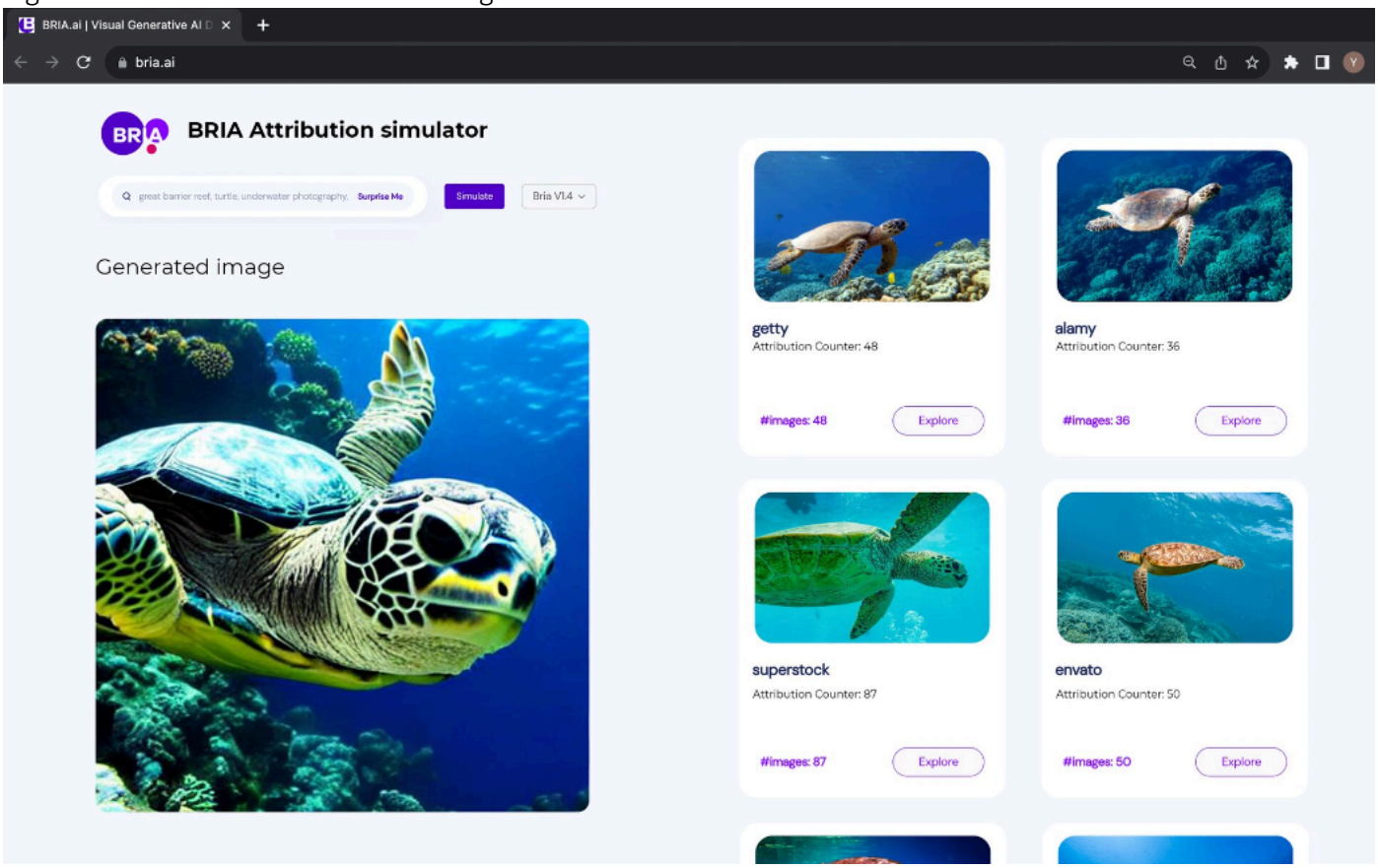


La vectorisation des images d'entraînement et des images générées dans un espace, latent ou prédéfini, d'indicateurs d'utilités, permet de calculer un score de ressemblance entre elles (figure 6). Cette technique permet en théorie de rémunérer chaque œuvre en fonction de sa contribution à chaque contenu généré. En revanche, rien ne garantit que la ressemblance n'est pas fortuite et qu'elle reflète une contribution réelle dans la construction du modèle d'IA.

Pour diluer ce biais dans un échantillon plus large, cette technique peut être élargie à deux ensembles d'images d'entraînement et générées. Les chercheurs d'Adobe calculent ainsi la contribution du premier ensemble, le dataset, au second (figure 7)[22].

Cette technique est commercialisée. La startup israelo-américaine Bria calcule ainsi la valeur d'utilité des œuvres originales pour chaque contenu synthétique[23], et la regroupe par dataset. Cette méthode n'est pas encore appliquée aux datasets publics, bien que rien ne semble s'y opposer en théorie.

Figure 8 – Score d'attribution des images d'entraînement de Bria[24]



On notera que la « similarité » des vecteurs est une notion scientifique détachée de toute notion juridique de « contrefaçon », de « ressemblance » ou d'atteinte à une œuvre, définie en droit français comme « toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi » (CPI, art. L. 335-3).

Il n'y a pas nécessairement contrefaçon ou atteinte par la reproduction d'un élément d'une œuvre, une « contribution » en l'occurrence. De la même manière lorsque qu'un auteur crée, un élément peut entrer dans sa composition, sans qu'il n'y ait contrefaçon ou atteinte à un tiers. La rémunération qui découle de la « similarité » ne devrait donc pas paralyser l'action de l'auteur en cas de contrefaçon.

Table 1 – 22 datasets of The Pile

Component	Raw Size	Weight
Pile-CC	227.12 GiB	18.11%
PubMed Central	90.27 GiB	14.40%
Books3†	100.96 GiB	12.07%
OpenWebText2	62.77 GiB	10.01%
ArXiv	56.21 GiB	8.96%
Github	95.16 GiB	7.59%
FreeLaw	51.15 GiB	6.12%
Stack Exchange	32.20 GiB	5.13%
USPTO Backgrounds	22.90 GiB	3.65%
PubMed Abstracts	19.26 GiB	3.07%
Gutenberg (PG-19)†	10.88 GiB	2.17%
OpenSubtitles†	12.98 GiB	1.55%
Wikipedia (en)†	6.38 GiB	1.53%
DM Mathematics†	7.75 GiB	1.24%
Ubuntu IRC	5.52 GiB	0.88%
BookCorpus2	6.30 GiB	0.75%
EuroParl†	4.59 GiB	0.73%
HackerNews	3.90 GiB	0.62%
YoutubeSubtitles	3.73 GiB	0.60%
PhilPapers	2.38 GiB	0.38%
NIH ExPorter	1.89 GiB	0.30%
Enron Emails†	0.88 GiB	0.14%
The Pile	**825.18 GiB**	

Valeur Trak

Aleksander Madry, professeur au MIT, est en charge du raisonnement de l'AI chez OpenAI. Il propose la solution *Trak*, fondée sur un algorithme approximatif du modèle d'AI pour en décrire l'évolution pendant l'apprentissage.

Il démontre que sa méthode, qui ne nécessite qu'un nombre limité d'entraînements, est économiquement adaptée aux plus grands modèles visuel+texte (CLIP) et de langage (BERT et mT5)[25].

Valeur approchante de Shapley

Boaz Barak, professeur de Harvard en charge de la sécurité (« superalignment »[26]), également chez OpenAI, propose quant à lui de mettre en œuvre la valeur de Shapley.

Son équipe de chercheurs montre en grandeur nature comment calculer la valeur de Shapley de datasets comme Wikipedia ou Github (table 1) pour un modèle d'IA à usage général formé à partir d'un ensemble[27] comparable à celui de GPT (table 2).

Selon elle, aligner la rémunération sur les contributions quantifiables de chaque dataset garantit la meilleure équité de la répartition des redevances[28]. Elle indique sa préférence pour la variante *In-run data Shapley*[29], moins onéreuse (figure 9).

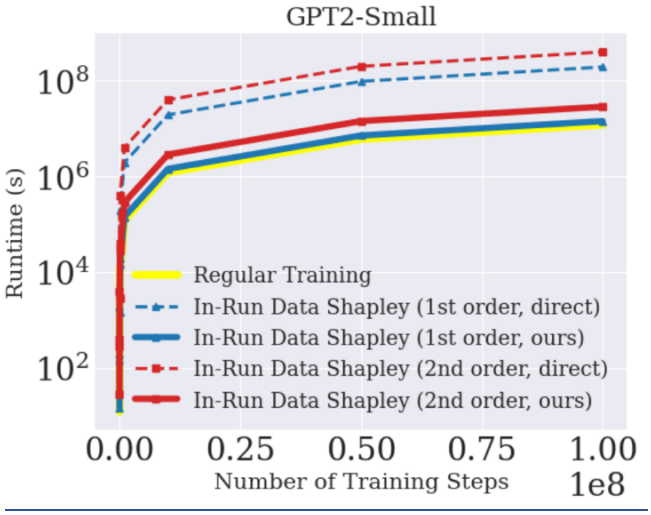
Table 2 – Datasets de GPT-3[30]

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

et GPT-4[31]

Dataset	Type	Fournisseur
Project Gutenberg	Livres	Public
Wikipedia	Encyclopédies	Public
BookCorpus	Livres	Public[32]
Google Books Ngrams	Livres	Google
WebText	Web	OpenAI
CommonCrawl	Web	Public
arXiv	Scientifique	Public
PubMed	Scientifique	Public
StackOverflow	Forums	Prosus
Reddit	Conversations	Reddit
CommonVoice	Multilingue	Public
GitHub	Code	Microsoft

Figure 9 – Consommation marginale de *In-Run data Shapley* par rapport à l'entraînement seul[33]



9 – Valorisation de l'œuvre dans le dataset

AI-generated by DALL-

Les banques de media (ex. Shutterstock, Getty, Audiosparx) vendent des datasets aux fournisseurs d'IA générative pour un prix confidentiel, mais rémunèrent les ayant-droits en fonction de leur volume de contribution à ces datasets.

Ce modèle « pay-to-train » établit une équité partielle entre auteurs :

- un auteur ayant contribué avec 200 photos reçoit, toutes choses égales par ailleurs, une rémunération deux fois supérieure à un contributeur ayant contribué avec 100 photos.

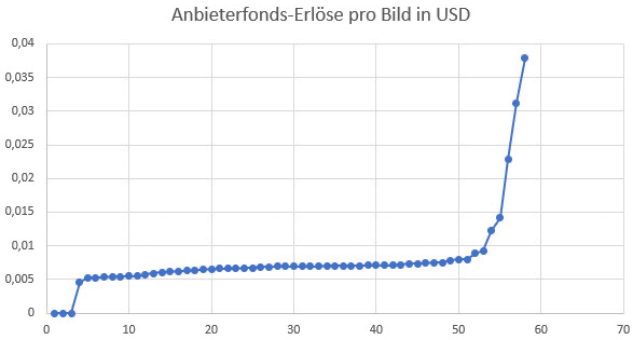
- en revanche la qualité n'est pas prise en compte, ni l'utilité pour le contenu généré. Les œuvres de Van Gogh sont en effet plus utiles que celles de Picasso si les utilisateurs insèrent plus souvent « à la manière de Van Gogh » dans leurs prompts.

L'avantage de ce modèle est sa simplicité de calcul. Il est opérationnel depuis deux ans[34], et semble bien accepté par les ayant-droits malgré son imperfection. Il nécessite néanmoins d'identifier les ayant-droits.

La valeur « pay-to-train » est calculable à partir d'indicateurs d'utilités divers (table 4). Shutterstock rémunère par exemple chaque photo à 0,0050 $ par an, auquel s'ajoute jusqu'à 0,0038 $ par photo et par an en fonction de la quantité de métadonnées et autres indicateurs[35] (figure 10). L'information de similarité de

chaque contribution au contenu généré, calculée dans la première étape (valeur du dataset), pourrait aussi être prise en compte pour affiner la valeur « pay-to-train ».

Figure 10 – Valeur Pay-to-train : rémunération des auteurs par Shutterstock. Echantillon N = 58 (en $/image)[36]



10 – Attribution de l'œuvre dans le contenu généré

La similarité des vecteurs dans un espace latent permet théoriquement, on l'a vu, d'attribuer une œuvre à un contenu généré, si on accepte un certain taux de fausses attributions.

De nouvelles méthodes sont en cours d'évaluation pour baisser ce taux.

Ainsi, les chercheurs d'Adobe proposent d'attribuer des données pour les modèles texte-image, en désapprenant l'image synthétisée et en identifiant les images de formation oubliées[37] (figure 11), ou de trouver des traces de tatouages des données d'entraînement dans les données générées[38] (figure 12).

Figure 11 - En modifiant le modèle pré-entraîné pour désapprendre le résultat synthétisé, le modèle oublie également les images d'entraînement influentes

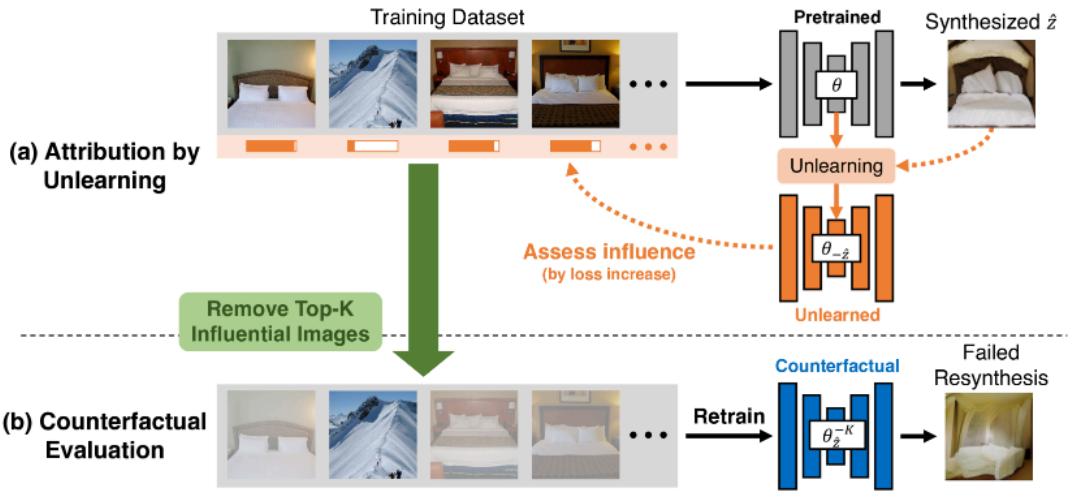
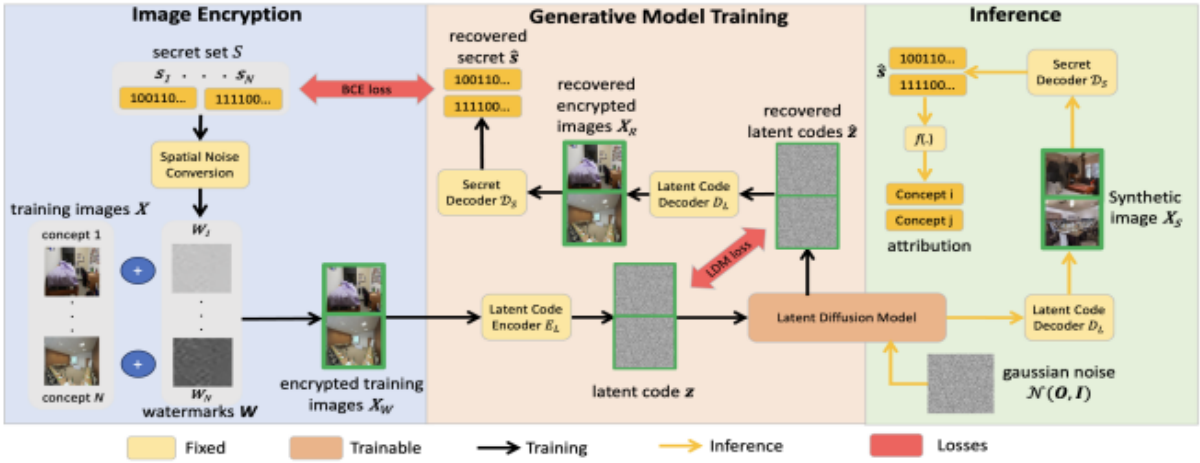


Figure 12 - Le modèle génératif est entraîné avec les images tatouées. A partir du tatouage extrait des images générées, l'algorithme est capable d'attribuer quelles images du jeu de données d'entraînement ont influencé la génération



11 – Identification des ayant-droits

AI-generated by Adobe Firefly

Les datasets publics donnent des éléments d'identification de qualité variable, qui peuvent ou non être recoupés, de manière plus ou moins facile, pour identifier chaque ayant-droits.

Les méthodes de recoupement sont surtout connues pour leur utilisation frauduleuse contre la vie privée. Dès 1997, la déclaration obligatoire de statistiques agrégées par un assureur du Massachussetts, avec des données électorales, avait permis d'identifier sans erreur possible le Gouverneur William Weld et de connaître son bilan de santé[39].

Leur utilisation légale faciliterait la ré-identification des auteurs : par comparaison de graphes sociaux, par corrélation statistique grâce à une combinaison rare d'attributs (niche scientifique, heures d'activité, adresse IP), par reconstruction d'attributs individuels à partir de moyennes de groupe et d'attributs connus d'autres individus, etc[40].

Table 3 – Identifiants des auteurs de datasets utilisés par GPT-4

	Dataset	Elément d'identification	Recoupement direct	Identifiant
Identification directe	Project Gutenberg	-	-	Nom, prénom, dates_naissance_et_mort
	arXiv	-	-	Nom, Initiale_prénom
	PubMed[41]	-	-	Id_author
Identification indirecte	CommonCrawl	Domaine	Whois	Registrant_contact (name, organization, street, city, state, phone)
	BookCorpus	Plain_text[42]	Smashwords	Nom, prénom
	WebText	Plain_text + search = « written by »[43]	API Amazon	Nom, prénom
	StackOverflow	HTML[44]	Stackoverflow.com	User_ID[45], network_profile[46]
En attente de revendication	Wikipedia	Pseudo_utilisateur	-	-
	Reddit	Pseudo_utilisateur	-	-
Autre	CommonVoice[47]	(stats)[48]	-	-
	Google Books Ngrams	(non applicable)	(na)	(na)
	GitHub	(non accessible)		

Pour augmenter le taux d'attribution, les Codes des bonnes pratiques du Bureau de l'IA et du Comité de l'IA (« Codes ») devraient indiquer les méthodes permettant

- de révéler le meilleur identifiant des ayant-droits dans le cadre de la réglementation de la protection de la vie privée
- de définir la norme de l'identifiant à appliquer (ex. ISWC[49] pour les œuvres musicales, ISAN pour les films[50])
- de gérer les revendications de droits[51] et les droits non attribués. Au même titre que les auteurs ont revendiqué leurs droits dès qu'ils ont pris conscience de la valeur commerciale du *Huffington Post* à laquelle ils avaient contribué[52], il est probable que les datasets issus de l'économie collaborative comme Wikipedia feront l'objet de revendications massives
- p.m : de grouper les ayant-droits représentant peu d'œuvres

12 – Optimisation

Toutes les solutions évoquées sont réglables et complémentaires :

- Les datasets font l'objet de traitements préliminaires. Les contenus de Wikipedia ont été expurgés des listes et tables pour l'entraînement de BERT[53]. A l'inverse certains datasets synthétiques sont constitués de toutes pièces pour une destination particulière[54].

- Les indicateurs d'utilité, qui sont les variables des algorithmes de rémunération, peuvent être choisis dans une liste étendue (table 4).

Table 4 – Usefulness indicators

researcher, company	Media	Usefulness indicators
Piao, Chen et al. [55]	Image	Feature vectors
Wang, Deng et al. [56]	Encyclopedia, Academic, Image, Code	Counterfactual model (S) Generated contents by the model $x^{(gen)}$ and the counterfactual model
Wang, Mittal et al. [57]	Encyclopedia, Academic, Image, Code, Email, Forum	Set of hold-out validation data points $Z^{(val)}$
Deng, Zhang et al.[58]	Music	Average velocity of events Average pitch height of all event Sum of the time deltas of all event
Shutterstock + Bria	Image, Vector image	Similarity score Turnover per dataset Data volume per dataset and per author Metadata volume Role played in the development of the original models Role played through royalty payments tied to future generative licensing activity
Adobe Stock + Bria	Image, Vector Image	Similarity score Image eligibility Image date Number of downloads

- Le mode de calcul de la valeur de Shapley est réglable
  - avec un nombre arbitraire de contenus générés, c'est-à-dire une taille d'échantillon
  - pour tout dataset, sur-dataset ou sous-dataset, jusqu'à chaque donnée individuelle, ce qui permet d'en régler la granularité(figure 13)
  - avec des fonctions approchantes[59], qui concèdent de la précision en échange d'un gain substantiel de temps de calcul (figure 14)
  - avec des nuances de définitions de la fonction d'utilité et des indicateurs d'utilité (table 4)

Figure 13 – Sous dataset de Commoncrawl, utilisé par GPT 3 (voir table 2)

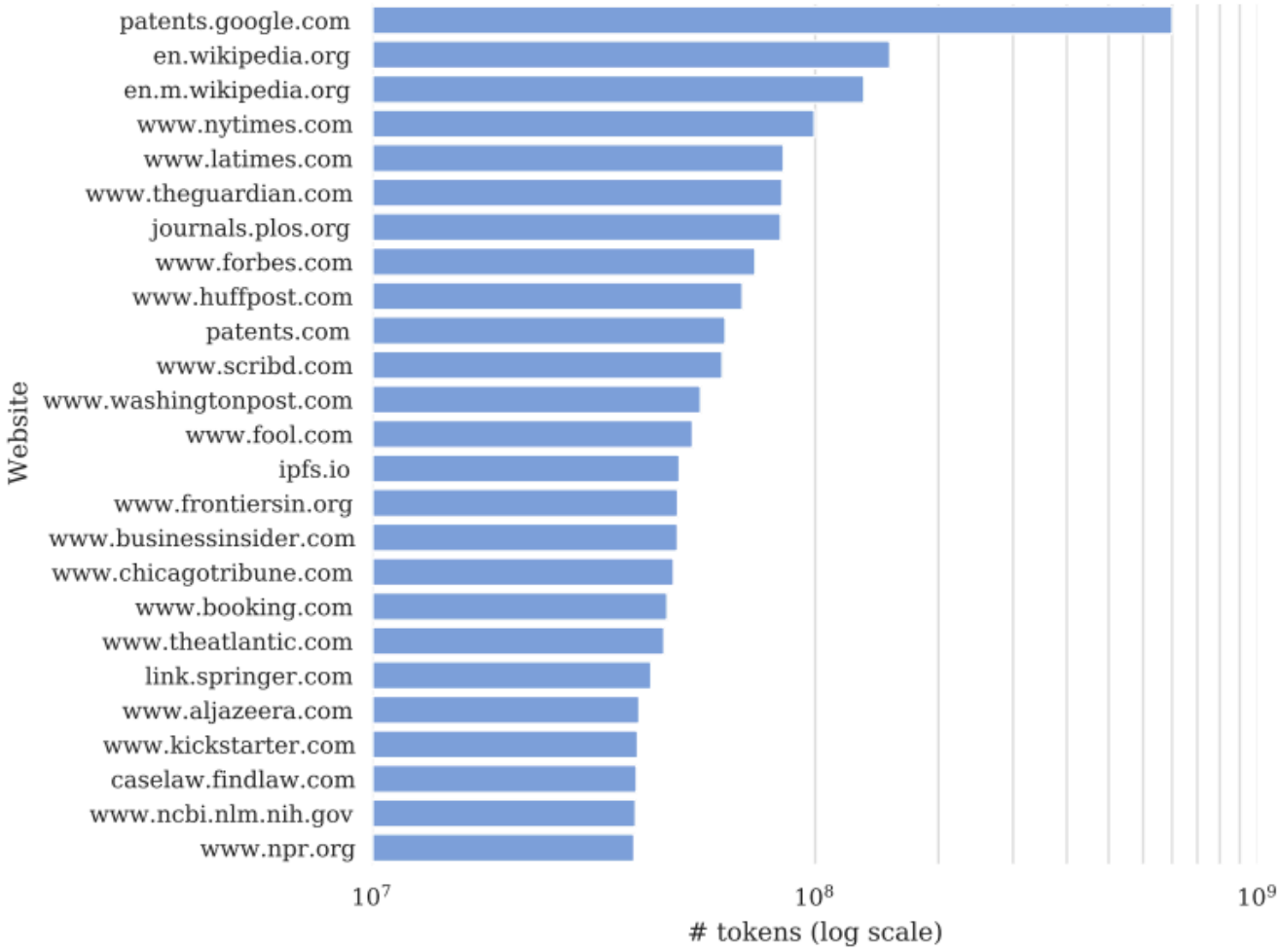
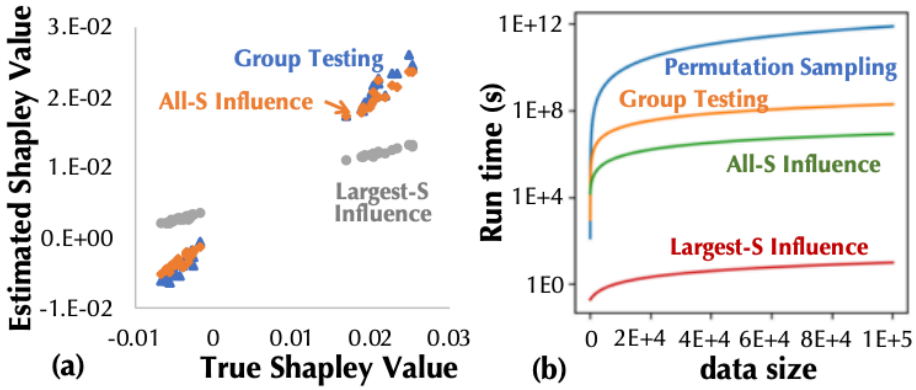


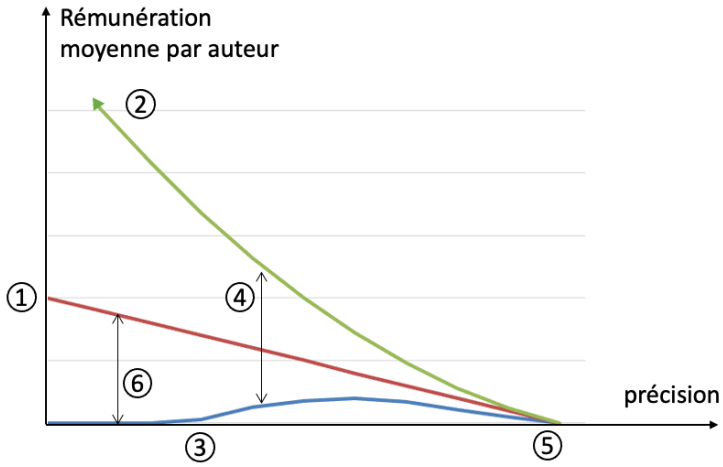
Figure 14 - Trois fonctions approchantes de la valeur de Shapley : perte de précision contre gain de temps de calcul[60]



- La précision de l'équité nécessite un calcul dont le coût pèse sur la rémunération moyenne par auteur, à collecte constante (figure 15).

Figure 15 – Impact du coût de la précision sur la rémunération.

(1) A précision nulle, la rémunération par auteur est égale à la collecte totale divisée par le nombre d'auteurs.
(2) A précision nulle, la rémunération potentielle maximale d'un auteur est égale à la collecte totale
(3) A précision nulle, la rémunération potentielle minimale d'un auteur est égale à zéro
(4) Ecart d'incertitude
(5) Quand le coût de la précision dépasse la collecte totale, la rémunération et l'écart d'incertitude sont nuls
(6) A collecte constante, la rémunération moyenne par auteur diminue avec la précision



- Les catégories d'ayant-droits ont des revendications variables. L'Authors'Guild recommande par exemple que les plateformes intègrent différentes manières de fixer les prix, que les auteurs pourront ensuite choisir ou refuser. De baser les tarifs sur l'économétrie et à partir de tous les revenus générés, en prévoyant les commissions des plateformes de licence, des agents et éditeurs[61].

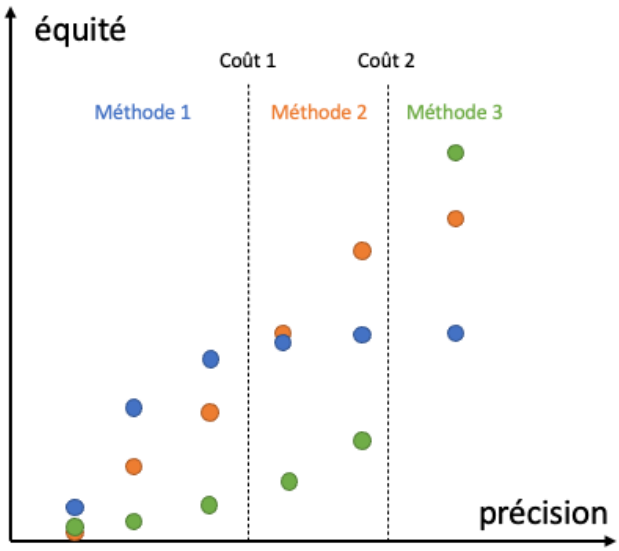
Compte-tenu de cette gamme de leviers de réglages, une solution optimale par cas d'usage devrait favoriser la meilleure équité de la rémunération.

Les Codes devraient encourager, au fur et à mesure des progrès techniques et sur la base d'un observatoire des bonnes pratiques, l'exploration de différents mixtes

- méthodes de constitution des datasets
- méthodes de pondération des indicateurs d'utilité (granularité, échantillonnage, volume, métadonnées)
- choix de fonctions d'utilité
- méthodes de calcul

pour maximiser l'équité de la rémunération à coût de calcul prédéfini (figure 16). Ces codes pourraient s'inspirer de ceux relatifs à la désinformation[62].

Figure 16 – Choix de la méthode : maximiser l'équité à coût de calcul maximum prédéfini



Mise en œuvre économique

13 – Méthode

Les Echos

IA générative : le débat politique est en retard d'une vague
18 janvier 2024

Le gouvernement français se retrouve à défendre l'innovation contre la culture, ce qui contredit sa vocation historique pour l'exception culturelle. Cette situation ne peut être que préjudiciable à la France et à l'Europe. Inédite, elle révèle en fait un impensé profond, celui du point d'équilibre désirable entre les intelligences humaine et artificielle. Lire

Si la situation actuelle persiste, il faudra attendre les décisions des juges, ce qui fera perdre 10 à 15 ans aux dépens :

- des auteurs, puisque dans l'intervalle ils ne perçoivent pas ou peu de rémunération
- des fournisseurs de modèles d'IA générative, qui se privent de données de qualité pourtant nécessaires au bon fonctionnement de leurs produits
- de l'Europe, qui perd une opportunité en tant que puissance régulatrice et visionnaire de l'économie de la créativité
- de la France, qui perd une opportunité en tant qu'héritière de Beaumarchais et porte-étendard de l'exception culturelle
- de l'intérêt général, pour lequel il faut apaiser les tensions qui s'opposent à l'efficience économique
- de l'humanité, dont le travail futur se dessine aujourd'hui dans le rapport entre intelligences humaine et artificielle

Comme l'entrelac des contraintes techniques, économiques et juridiques est complexe, nous proposons une méthode par étapes, détaillée dans les paragraphes suivants : 1/ évaluer l'acceptabilité du pay-to-train ; 2/ évaluer l'acceptabilité de la valorisation des datasets ; 3/ illustrer l'opportunité commune par les enjeux économiques ; 4/ proposer une vision de l'intérêt général ; 5/ y inscrire le principe de la rémunération équitable ; 6/ décrire les étapes de la mise en œuvre de la solution Controv3rse ; 7/ la soumettre à la critique des parties prenantes et l'améliorer.

14 – Acceptabilité du pay-to-train

JDN
JOURNAL DU NET

#FairlyShare contre le péché originel d'Internet – septembre 2014, par Vincent Lorphelin, Gilles Babinet et al.

Ce qu'ils dénoncent ? Les plate-formes qui reposent sur l'industrialisation du travail gratuit, en dehors des règles de l'appel d'offre, du concours ou du bénévolat. Ils emboîtent ainsi le pas à la campagne NO!SPEC.com, qui exhorte les designers à ne pas succomber aux sirènes des plate-formes comme crowdspring ou 99designs. Ou encore à la levée de boucliers qui a suivi le rachat du journal de crowdsourcing Huffington Post par AOL pour 315 millions de dollars. Lire

La méthode « pay-to-train » est actuellement la plus répandue pour rémunérer les auteurs dont les œuvres sont utilisées pour l'entraînement des modèles d'IA à usage général (table 5). Malgré une transparence et une équité partielles, elle n'a pas suscité de procès ni de pétitions, comme on en avait vu par exemple lors de la cession du Huffington Post (voir encadré).

Table 5 – Banques de media et modèles d'IA utilisant leurs datasets

DATE	MEDIA	FOURNISSEUR DE DATASET	FOURNISSEUR DE MODELE D'IA
10/25/22	Image, photo	Shutterstock	OpenAI
01/12/23	Image, photo	Shutterstock	Meta, Google, Amazon, Apple
03/12/23	Image, photo	Shutterstock	LG (Exaone)
03/21/23	Image, photo	Getty Images	Nvidia
05/29/23	Avatar	Ascendant Art	
07/11/23	Image, photo	Shutterstock	OpenAI
08/08/23	Image, photo	Shutterstock	Nvidia (Picasso)
09/07/23	Image, photo	Getty Images Alamy Bria	
09/13/23	Music	AudioSparx	Stability AI
09/13/23	Image, photo	Adobe Stock (Firefly)	
09/26/23	Image, photo	Getty Images (Getty Gen AI)	
10/04/23	Image, photo	Canva (Magic Studio)	Open AI (Dall-e) Google (Imagen)
02/21/24	Image, photo	Superstock[63] Envato	
04/11/24	Image, photo	EyeEm Freepick Photobucket	
06/27/24	Music, Voice	Universal Music* Sony Entertainement Warner Records	Google (Youtube)
07/30/24	3D	Shutterstock[64]	Nvidia (Edify)

*expérimentation

Les contrats d'achats de datasets auprès des banques de media ont plutôt été accompagnés de discours qui valorisent la rémunération « respectueuse, responsable, sécurisée et équitable » des auteurs (table 6).

Table 6 – Fairness of pay-to-train

PLATFORM	STOCK AND DATASET	CONSUMER DEMAND
Nvidia	**Getty, Shutterstock, Adobe**	Nvidia : "We created a platform that makes it possible for our partners to train from data that was licensed properly from, for example, Getty, Shutterstock, Adobe," Huang said. "They're respectful of the content owners. The training data comes from that source, and whatever economic benefits come from, that could accrete back to the creators."[65]
	Adobe	Adobe : we are developing generative AI responsibly, with creators at the center[66]. Adobe : Create with confidence knowing your output is the highest quality, built with models that are designed to be safe for commercial use[67].
	Getty	Getty : today announced the launch of Generative AI by iStock, an affordable and commercially safe generative AI tool [..]without fear that something that is legally protected has snuck into the dataset and could end up in their work[68]
Meta	**Shutterstock**	Shutterstock : expertise in creating a scaled ecosystem that compensates and connects contributors to creators[69].
LG Exaone	**Shutterstock**	Shutterstock : AI-generated content ownership cannot be assigned to an individual and must instead compensate the many artists who were involved in the creation of each new piece of content[70]. Our customers can safely and legally license AI images for their own purposes, worry free. We also pay contributors whose works train our models, so you can use our AI with a clean conscience, as well[71]. Our AI image generator remains the safe, easy to use, responsibly built feature you know and love[72] LG : AI model is being trained with millions of high-resolution images and metadata from Shutterstock and will convert text-based prompts into images. Contributors whose works were used to train the model will be compensated via Shutterstock's Contributor Fund and will also be compensated whenever new generative content that uses their IP is created and licensed by customers[73].
Open AI	**Shutterstock**	OpenAI : E The data we licensed from Shutterstock was critical to the training of DALL-E[74]. Ensuring that the creator economy continues to be vibrant is an important priority for OpenAI. Writers, artists, composers and other creators have contributed immeasurably to societies throughout the history of civilization [..] OpenAI does not want to replace creators[75].
	Canva	Canva : Magic Media's text-to-image, DALL·E by OpenAI, and Imagen by Google Cloud […]ensure their AI models are fair and ethical[76]. "we share the value that we're developing in Canva with the creators who have been with us all these years. With the new Creator Fund, they get access to a royalty pool that is aligned with AI creations."[77]
Google Cloud	**Canva** **Shutterstock**	Google Cloud : Shutterstock has also emerged as a leading innovator being the first to launch an ethically-sourced AI image generator, now enhanced with Imagen on Vertex AI.
Google Youtube	**UMG**	Google's YouTube is trying to make AI music deals with music majors[78]. He wants to ensure "music rightsholders get paid for their training data contributions[79]. UMG and YouTube are in talks to license artists' voices and melodies to train AI models[80]".
Amazon SageMaker	**Alamy** **Getty** **Envato**	Alamy : Central to the ethos of The Fair Diffusion Program is a commitment to fairly compensate photographers, artists and creators[81]. Getty : we hold to the highest ethical standards and respecting of the intellectual property and personal privacy rights of others[82].

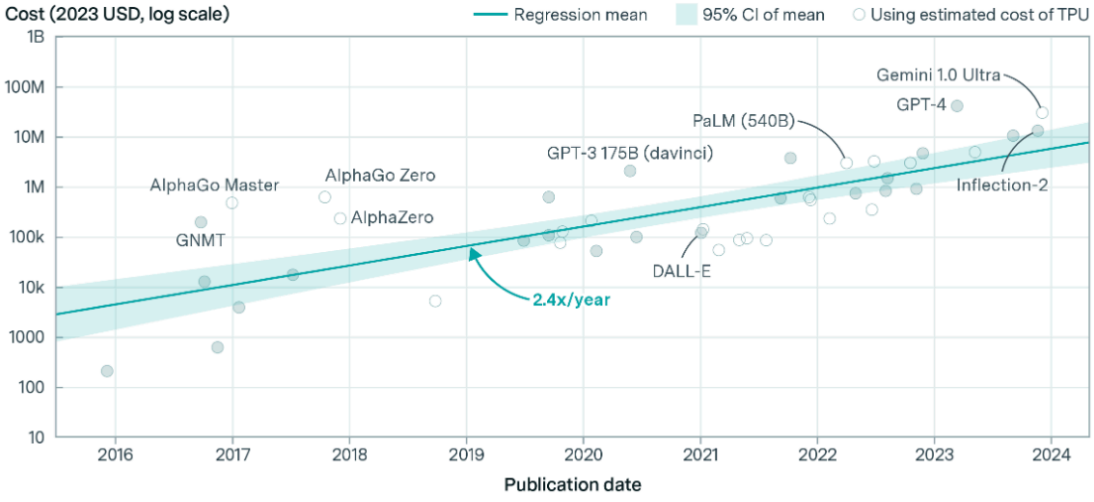
PLATFORM	STOCK AND DATASET	CONSUMER DEMAND
Sony		Sony's President says: "New products and businesses built with music must be developed with the consent of the owner and appropriate compensation and credit. It is essential to understand why the training of AI models is being done, what products will be developed as a result, and what the business model is that will monetize the use of the artist's work"[83]. "If a generative AI model is trained on music for the purpose of creating new musical works that compete in the music market, then the training is not a fair use. Training in that case, cannot be without consent, credit and compensation to the artists and rightsholders."[84]
Mistral AI		Arthur Mensch: **"**we have an interest in having access to quality content. We therefore began discussions with content providers, in publishing and in the press. There are synergies and value sharing mechanisms to be found."[85]
Midjourney		"There isn't really a way to get a hundred million images and know where they're coming from. It would be cool if images had metadata embedded in them about the copyright owner or something. But that's not a thing; there's not a registry. There's no way to find a picture on the internet, and then automatically trace it to an owner and then have any way of doing anything to authenticate it."[86]

15 – Acceptabilité de la valorisation des datasets

La valorisation par similarité vectorielle est commercialisée par Bria et semble bien acceptée.

La valorisation par la méthode de Shapley est plus délicate. Le calcul complet de la valeur de Shapley nécessite un temps de calcul largement supérieur à celui de l'entraînement même du modèle d'IA (figure 17), ce qui le rend économiquement dissuasif.

Figure 17 – Coûts d'entraînement des modèles d'IA à usage général[87]



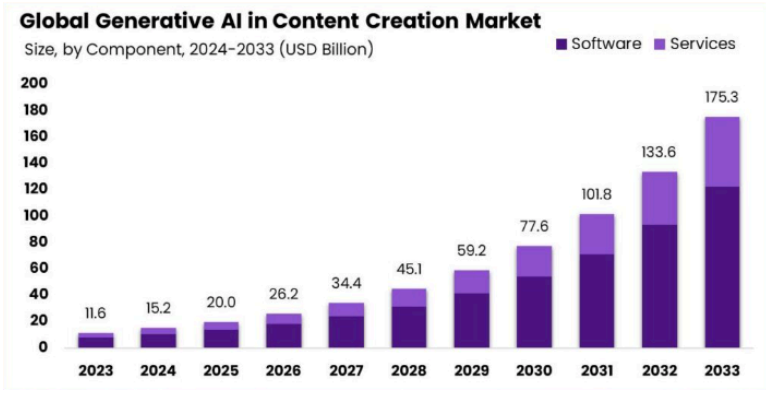
C’est pourquoi la sobriété d’autres méthodes approchantes a été évaluée, quitte à perdre en précision des valeurs calculées. *In-run data Shapley 1st-order*[88], en particulier, permettrait d’effectuer le calcul en même temps que l’entraînement normal et ne nécessite pas de puissance informatique supplémentaire.

Cette méthode a été proposée par des chercheurs d’Open AI et d’autres leaders technologiques.

Table 7 – Companies research for fair compensation models

COMPANY	RESEARCH TEAM	DATE	MOTIVATION
OPEN AI Princeton University Columbia University Harvard University University of Pennsylvania	**An Economic Solution to Copyright Challenges of Generative AI**[89] Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, Weijie J. Su	April 14, 2024 Sep 9, 2024 (v4)	Shapley values offer a principled approach to distributing gains depending on the utility. The utility offers a way to measure the extent to which the data sources are responsible for generating the content. It is small if the counterfactual model is unlikely to generate the same content as the deployed model, and vice versa. The utility can be interpreted as the total compensation all members collectively deserve for providing their data to train the generative AI model.
ADOBE Carnegie Mellon University UC Berkeley	**Evaluating Data Attribution for Text-to-Image Models**[90] Sheng-Yu Wang, Alexei Efros, Jun-Yan Zhu, Richard Zhang	August 8, 2023	A method that fairly attributes the training images opens potential possibilities where creators can be incentivized and rewarded for providing data
	ProMark: Proactive Diffusion Watermarking for Causal Attribution[91] Vishal Asnani, John Colomosse et al.	March 14, 2024	[We] demonstrate rather than approximate/imply causation. This provides confidence in grounding downstream decisions such as legal attribution or payments to creators.
HUAWEI University of British Columbia Simon Fraser University KTH Royal Institute of Technology	**Improving Fairness for Data Valuation in Horizontal Federated Learning**[92] Zhenan Fan, Huang Fang, Zirui Zhou, Jian Pei, Michael P. Friedlander, Changxin Liu, Yong Zhang	May 23, 2022	The motivation of data owners partially depends on whether the collaboration and rewarding in federated learning are fair

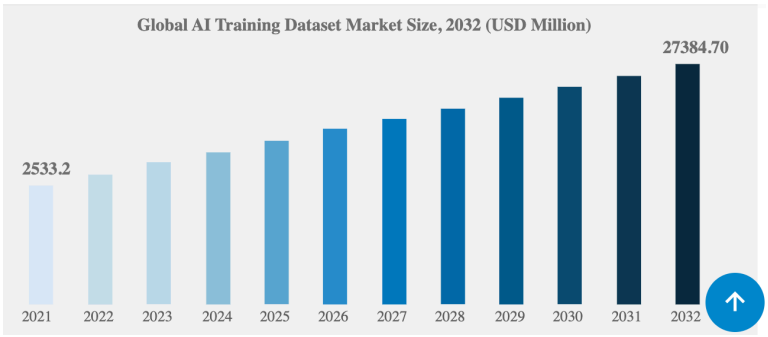
16 – Enjeux économiques



© Market.us 2024

Pour le seul segment de la créativité, la taille du marché mondial de l'IA générative est estimée à 11,6 milliards USD en 2023 et dépassera 100 milliards USD d'ici 2031[93].

Pour disposer d'ordres de grandeur, un taux de rémunération équitable de 15% représenterait davantage que la collecte actuelle des OGC[94] pour les droits artistiques (musique, audiovisuel, spectacle vivant, littérature et arts visuels).



© Business Research Insights 2024

Les entreprises spécialisées dans les datasets de formation à l'IA jouent un rôle central dans la collecte, l'annotation, la conservation des données et leur adéquation aux besoins des algorithmes d'apprentissage.

Le marché mondial des datasets de formation en IA était de 2,5 milliards de dollars en 2021 et devrait atteindre 27 milliards de dollars d'ici 2032[95].

© Business Research Insights 2024

Les datasets se classent selon six modèles d'affaires principaux :

- Scraping : DALL·E, par exemple, a d'abord été formé avec des centaines de millions d'images sous-titrées provenant d'Internet[96]. Cette pratique est à l'origine de nombreux procès[97].

- Uber des données : les créateurs de modèles d'IA ont commencé à couvrir les risques et à sécuriser les chaînes d'approvisionnement en données, à la fois via une industrie florissante de courtiers en données et par des accords avec les propriétaires de contenu qui ont surgi pour satisfaire la demande.

Une industrie d'entreprises de données d'IA spécialisées est en train d'émerger, créant des réseaux de travailleurs sous contrat à court terme pour produire des visuels personnalisés et des échantillons de voix à partir de zéro, ce qui s'apparente à une économie de travail de type Uber pour les données[98].

Les prix sont de 1 à 2 dollars par image, entre 2 et 4 dollars par vidéo courte et entre 100 et 300 dollars par heure pour les films plus longs. Le tarif du marché pour le texte est de 0,001 dollar par mot. Les propriétaires des photos, des podcasts et des données médicales sont rémunérés à hauteur d'environ 20 à 30 % du montant total de la transaction.

- Banques de contenus : Shutterstock, Getty, Canva ou Adobe disposent de grandes bases de media alimentées par des auteurs, rémunérés en fonction des datasets vendus. Shutterstock, leader pour les images, a vendu des licences aux fournisseurs d'IA pour plus de 100 millions de dollars, soit 12% de son chiffre d'affaires total, et devrait augmenter de 40% en 2024[99]. Ces licences d'images, video et musique formeront un marché adressable de 10 milliards de dollars en 2030, en croissance annuelle de 22% sur la période[100]. Nous estimons, en tendance, que la rémunération des auteurs s'établira en-dessous de 4% du marché des contenus générés, et baissera progressivement (table 8).

Table 8 – Author royalties market

$B	2023	2024	2025	2026	2027	2028	2029	2030	2031
Generative AI in Content Creation Market	11.6	15.2	20.0	26.2	24.4	45.1	59.2	77.6	101.8
Generative AI Training Dataset Market	2.3	3.1	3.8	4.6	5.6	6.8	8.3	10.1	12.4
Author royalties	0.5	0.6	0.8	0.9	1.1	1.4	1.7	2.0	2.5
%Content Creation Market	4%	4%	4%	4%	3%	3%	3%	3%	2%

- Groupes de media : les propriétaires de bases de données cèdent des licences sur les datasets, sans rémunération complémentaire pour les auteurs (table 9)

Table 9 – Media groups deals

DATE	CONTENT	MEDIA GROUPS MODEL
07/13/23	Text	Associated Press
12/13/23	Text	Axel Springer
04/29/24	Text	Financial Times
05/06/24	Text	Stack Overflow
05/16/24	Code	Reddit

- Sociétés d'auteurs : the Authors'Guild veut[101], grâce à la plateforme *Created by Humans,* permettre aux auteurs d'apporter leurs œuvres et de définir les options de leur licence pour les IA génératives[102].

D'autres initiatives similaires fleurissent comme *Human Native AI*[103], *Fairly Trained*[104] ou *Dataset Providers Alliance*[105]. La Commission de l'IA, en France, a proposé une plate-forme publique de mise en relation des fournisseurs de données et des fabricants de modèles, que décrit Alexandra Bensamoun : « l'idée est de créer une infrastructure facilitatrice. Cela nécessiterait une volonté politique et pourrait être porté par la Bibliothèque nationale de France ou l'Institut national de l'audiovisuel, dans le respect des droits. Ces plates-formes disposent de contenus, car la presse,

l'édition ou l'audiovisuel les leur envoient, au nom du dépôt légal. Mais elles n'ont pas le droit, sur ce fondement, de les mettre à disposition, sauf s'ils sont tombés dans le domaine public. Il faudrait donc que les titulaires de droits leur demandent de servir d'intermédiaire avec les fabricants d'IA, à tel prix, telles conditions… La rémunération leur serait ensuite redistribuée ».

La répartition du marché entre ces segments n'est pas claire. Le scraping représente la majorité des pratiques mais probablement une faible partie du marché. Le segment Uber des données devrait être le plus dynamique tandis que le modèle des sociétés d'auteur est à peine en phase de test.

Le marché potentiel est supérieur à celui des OGC, et représente une opportunité pour les ayant-droits, qui pourraient en retirer des revenus additionnels significatifs. Symétriquement les exploitants d'IA générative pourraient accéder à des bases de contenus de qualité. Ce potentiel, qui pour l'instant bénéficie aux segments alternatifs du marché, sera libéré par la mise en œuvre d'une solution sacémisante.

17 – Intérêt général

Interview de Vincent Lorphelin

Le Soir (Bruxelles), Janvier 2024

« Sam Altman et Elon Musk disent que l'IA générative va rendre le travail obsolète, il faut donc se préparer au revenu universel. Cela coagule avec la thèse de la fin du travail relayée par Jeremy Rifkin il y a quelques années. Cela rejoint aussi celle de la singularité qui avance que d'ici quelques années, l'IA générale va dépasser l'humain dans tous les domaines. La preuve c'est que l'on a toujours dit que l'IA n'ira jamais dans les métiers créatifs (elle ne sera jamais auteure, dessinateur...). Sauf que depuis l'an dernier, on y est. Les portes du bastion sont enfoncées.[…]

A l'inverse, l'ADN européen, c'est l'économie sociale de marché, c'est l'économie qui fait société. Le travail fait sens. Historiquement et d'un point de vue stratégique, la gestion collective des droits est deux fois plus développée en Europe qu'aux Etats-Unis. De ce fait, si on en revient à cette question fondamentale de l'intérêt général, pour faire contrepoids à la thèse de la singularité, l'intérêt général européen c'est de valoriser le modèle décentralisé.

Imaginons que ce modèle de plateforme décentralisée devienne systémique, fasse système, au même titre que le font, aujourd'hui, les plateformes centralisées comme les Gafam. Dans ce cas, la matière manipulée, ce sont des droits intellectuels, pas des marchandises. C'est de la valeur économique non marchande. L'économiste Karl Polanyi avait défendu l'idée que l'économie de marché a marchandisé le travail, alors que le travail ne peut pas être assimilé à une marchandise.

Les plateformes centralisées, elles, vont tenter au maximum de donner une valeur marchande aux droits intellectuels. Dans une logique décentralisée, les droits vont s'exprimer en pourcents, pas en monnaie. Les valeurs économiques manipulées ne s'expriment plus en euros ou en dollars. C'est une économie qui est fondée non pas sur de l'argent, mais sur une propriété. Pour faire un parallèle avec ce qu'Emile Zola avait appelé l'argent liquide (avec le passage d'une économie où la richesse ne venait plus de la propriété terrienne mais de la circulation de l'argent), on propose la notion de

Les Echos

IA: faut-il attribuer des droits humains aux machines ?
6 mai 2024

Brad Smith, président de Microsoft, est en train de commettre une faute morale contre la société. En demandant la suppression des droits d'auteur pour assurer le respect du « droit aux machines d'apprendre » il décide de promouvoir l'IA plutôt que l'intelligence humaine. Lire

« propriété liquide », où le fonds de roulement de l'économie c'est la circulation de droits de propriété intellectuelle.

Aujourd'hui, on a tellement tout financiarisé, qu'une entreprise présente des comptes comme si tous les jours elle était à vendre, parce qu'elle mesure tout en argent. Or, il y a des richesses immatérielles, hors-bilan, qui expriment des valeurs qui ne s'expriment pas en argent. La richesse de l'entreprise dépasse ce qu'exprime l'argent. Dès lors que l'on manipule des propriétés liquides, on est amené à trouver d'autres indicateurs.

Progressivement, on évolue vers une société qui solvabilise une richesse d'un autre ordre. D'accord, ce sont des visions d'économiste à long terme, mais on remarque que l'IA, de manière contre-intuitive, fait glisser les modèles économiques vers cette vision-là » (lire la suite ici).

La sacémisation ouvre ainsi, au-delà du seul segment de la créativité, la perspective de la valorisation du patrimoine humain et culturel par la propriété liquide. L'Union européenne s'était fixée à Lisbonne de « devenir l'économie de la connaissance la plus compétitive et la plus dynamique du monde ». Les nouvelles technologies disponibles permettent de relancer cette vision pour l'intérêt général européen.

Table 10 - Influential visionaries on digital ownership

NAME	COMPANY	CITATION
Chris Dixon	**a16z**	"The movement [of web 3] has the potential to bring back the spirit of the early internet; secure property rights for creators; reclaim **user ownership** and control; and break the stranglehold big, centralized companies have on our lives"[106].
Neal Stephenson	**Lamina1 (inventor of metaverse)**	"you can actually track the degree of influence that any given image or any given input had on the final result. And then if that final result is worth something, then maybe there's a way to give credit and some compensation in proportion to those inputs [..] Let's say that I were to write a book that had a magic sword that was just described in the book. [..] Somebody might then create an asset that could be sold on the Unreal asset store [..] You've got sound designers who need to do something similar with the sounds that it makes when it's used. And you've got programmers who need to, using blueprints or C++, who need to integrate the sword into the game so that it actually is capable of doing something and contributing to the experience. So at each stage, more value is being added. And at the end of that process, you've got something that might actually bring in some revenue. And when that revenue finally appears, what you'd like it to do is **propagate backwards**. And you'd like the different people who contributed to the value chain to get compensated in some way"[107]

NAME	COMPANY	CITATION
Jaron Lanier	**Microsoft**	"People will be paid for their data and will pay for services that require data from others. Individuals' attention will be guided by their self-defined interests rather than by manipulative platforms beholden to advertisers or other third parties. Platforms will receive higher-quality data with which to train their machine learning systems."[108] Lanier acknowledges that even **data-dignity** researchers can't agree on how to disentangle everything that AI models have absorbed or how detailed an accounting should be attempted. Still, Lanier thinks that it could be done — gradually[109]. "we have to calculate and present the provenance of which human sources were the most important to a given AI output. We don't currently do that. We can do it efficiently and effectively, it's just that we're not. It has to be a societal decision to shift to doing that"[110].
Yat Siu	**Animoca Brands**	"Data is the most valuable of resources. For starters, it's powering all of the AI that we've been hearing so much about. No data, no ChatGPT. **Data is the new labor**. And we're not being fairly compensated for it.[..] There is a clear correlation between property rights and the wealth of nations."[111] "Imagine if you went onto Facebook and, at the end of the day, it showed that you just made a thousand dollars for the company. Your relationship to Facebook would completely change. You would probably demand your fair share"[112].
Trip Adler	**former CEO of Scribd**	"the **Fourth Law** is a set of guiding principles for how AI companies can use and train on human-created content. Fourth Law, inspired by science fiction author Isaac Asimov's three laws of robots, states that humans should have the right to consent and control how AI uses their work and be compensated (if requested) and credited for their work (if a book is referenced in the output, there should be a link to buy it)."[113]

18 – Rémunération équitable

Le Monde

IA générative et droits d'auteur : « La culture artificielle ne doit pas suivre le triste chemin de la malbouffe »
15 septembre 2023

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 une comparaison plus établie, le taux de 15 % est celui que retiennent les pays producteurs de pétrole sur le prix du baril. Lire

Nous prenons le droit à rémunération, tel que collecté par l'ASCAP[114] et la Spré, comme hypothèse de cette note[115] (voir encadré).

Pour mémoire, la Spré prélève en France une redevance calculée sur le chiffre d'affaires des discothèques au bénéfice des musiciens. Un échantillon représentatif de 120 d'entre elles est équipé d'un boîtier « Yacast », financé par la Sacem et la Spré. Yacast collecte les listes de musiques diffusées, et publie des statistiques[116] par artiste, label, genre et nombre de semaines de diffusion. Après retenue statutaire de 9%, la Spré répartit le prélèvement aux OGC (SCPA 50%, ADAMI 25% et SPEDIDAM 25%)[117] et leur communique les statistiques Yacast pour rémunérer les ultimes bénéficiaires (artistes interprètes, producteurs, actions d'intérêt général).

Table 11 – Benchmark de la Spré

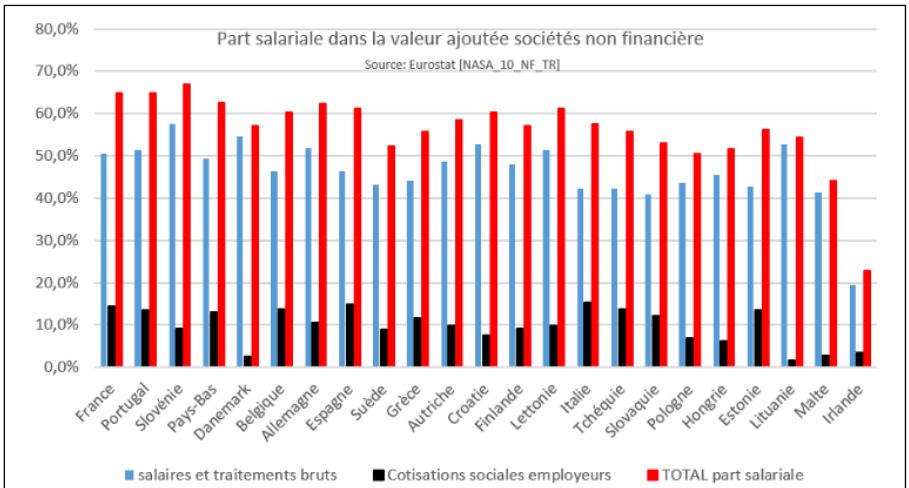
	Discothèque	Fournisseur de modèle d'IA à usage général
Fixation du taux de redevance	commission légale	commission légale[118]
Assiette du taux de redevance	chiffre d'affaires	chiffre d'affaires des contenus générés
Organisme de collecte	SPRÉ	à définir par le Bureau européen de l'IA
Échantillon	120 discothèques	N contenus générés
Statistiques	par artiste	par ayant-droit
Coût de gestion	retenue statutaire (9%)	forfait à exprimer en pourcentage de la redevance
Organismes de répartition	OGC (SCPA, ADAMI, SPEDIDAM)	OGC

Comment définir le taux équitable ?

Nous évaluons la part actuelle des auteurs à 4% du marché des contenus générés pour les industries créatives, en baisse progressive (table 8).

Dans son rapport de 2020, *L'auteur et l'acte de création,* Bruno Racine déplorait que « la rémunération proportionnelle des auteurs, n'atteint ou dépasse qu'exceptionnellement 10 % des recettes d'exploitation »[119]. Il attribue ce constat à « la relation qui lie l'artiste-auteur aux acteurs de l'aval (éditeurs, diffuseurs, producteurs) [qui] apparaît profondément déséquilibrée, ce qui conduit, le concernant, à mettre en cause dans de nombreux cas l'idée même de liberté contractuelle »[120]. Il identifie deux leviers pour y remédier :

- Celui d'un rééquilibrage des rapports de force dans la négociation. Il souhaite ainsi « renforcer les artistes-auteurs collectivement, par l'organisation rapide d'élections professionnelles qui permettront de donner corps et légitimité au Conseil national des artistes-auteurs à créer afin de servir de cadre à la négociation collective avec les diffuseurs ». La négociation autour de l'IA générative étant bloquée, ce rééquilibrage n'aurait pas d'effet.



© IFRAP 2022

- Celui de la valeur-travail, qui devrait fonder « un dispositif contractuel rémunérant le temps de travail de l'auteur, préempté par le commanditaire »

Comme l'IA générative est aussi gourmande en capitaux, l'approfondissement de cette approche réveillerait l'éternel débat de la juste répartition de la valeur entre capital et travail (voir encadré).

Nous proposons un troisième principe, celui de l'utilité. C'est sur ce principe qu'est conçue la valeur de Shapley et que l'économiste Ernst Fehr évalue à 14% la valeur apportée par la presse à Google Search[121]. Depuis la fameuse question « ce commentaire vous a-t-il été utile ? » lancée par Amazon[122], tous les sites web commerciaux améliorent la qualité de leur offre en multipliant les métriques de l'utilité, qui sont maintenant d'usage courant.

Le taux équitable devrait se définir comme la valeur de l'utilité de l'ensemble des contributions des auteurs rapportée au chiffre d'affaires des contenus générés. Celui-ci devrait être estimé par une société d'études de marché pour le Bureau de l'IA à partir des chiffres d'affaires des fournisseurs des modèles d'IA, et la collecte répartie au pro rata de ces chiffres d'affaires.

Fixer l'assiette sur le chiffre d'affaires du contenu généré permet à la rémunération équitable de l'auteur de ne pas dépendre de la plus ou moins grande segmentation verticale de la filière, par exemple : auteur / plateforme / dataset / modèle général / fine tuning / interface / génération.

Le Monde
Le triomphe de l'économie de l'utilité – juin 2015

L'opposition capital-travail a structuré la pensée économique pendant deux siècles. Bien que la théorie de l'utilité soit antérieure, elle a été supplantée par la théorie ricardo-marxiste de la valeur travail, puis par la théorie néoclassique de la valeur de marché. La transition numérique apporte de nouvelles métriques de l'économie, qui en changent la nature. Lire

Recommandations

19 – Étapes

Nous recommandons la mise en œuvre de la rémunération équitable selon les étapes suivantes :

Table 12 – Mise en œuvre de la rémunération équitable pour l'IA générative

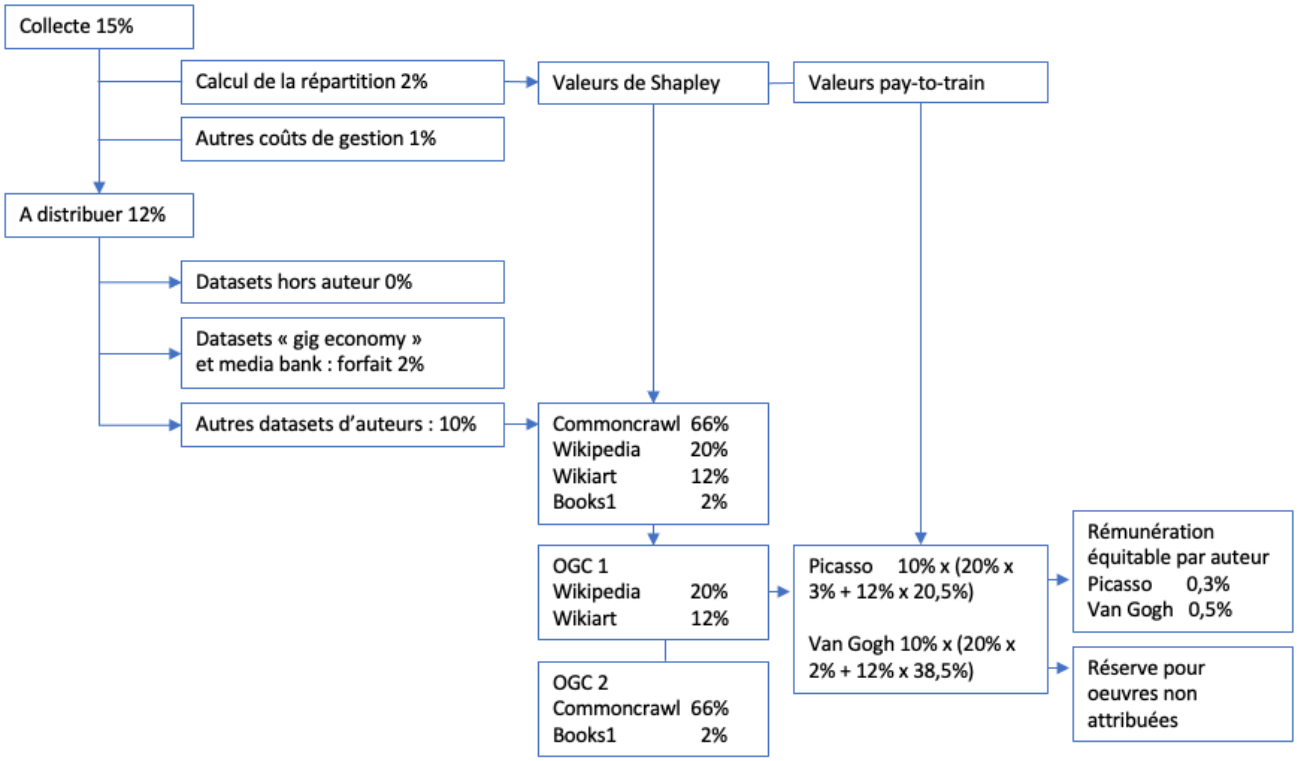
Recommandation n°	Processus	Bonnes pratiques, comparables
1 - Méthode	Solution « bêta » : discuter à partir d'une solution à parfaire pour faciliter la projection des interlocuteurs dans un résultat, et réduire la complexité du débat[123]	Méthodes de prototypage ou essai-erreur
2 - Vision	Décentralisation vs. plateformes centralisées, sacémisation vs. ubérisation, travail des auteurs vs. marchandisation, absurdité du « droit » des machines, propriété liquide, métriques de l'utilité	Cadre européen : économie sociale de marché, économie de la créativité, héritage et diversité culturelle
3 - Principe de la rémunération équitable	(Hypothèse de notre étude[124]) Mesure de l'utilité des contributions pour une attribution auteur par auteur. Affiner la précision de cette mesure avec les progrès techniques.	Statistiques Yacast pour la Spré
4 - Taux équitable	Définition : utilité de l'ensemble des œuvres d'auteurs pour l'ensemble des contenus générés Assiette : chiffre d'affaires des contenus générés Fixation : commission légale	Directive 2006/115/CE (13) : La rémunération équitable « devrait tenir compte de l'importance de la contribution apportée au phonogramme et au film par les auteurs et les artistes interprètes ou exécutants concernés »[125] Droits voisins (73) : « La rémunération des auteurs et artistes interprètes ou exécutants devrait être appropriée et proportionnelle à la valeur économique réelle ou potentielle des droits octroyés sous licence ou transférés, compte tenu de la contribution de l'auteur ou de l'artiste interprète ou exécutant à l'ensemble de l'œuvre ou autre objet protégé

		et de toutes les autres circonstances de l'espèce, telles que les pratiques de marché ou l'exploitation réelle de l'œuvre. Un montant forfaitaire peut également constituer une rémunération proportionnelle, mais cela ne devrait pas être la règle.[126] »
5 - Résumé suffisamment détaillé	Contenu public : liste des datasets, descriptif des méthodes choisies Contenu secret[127] : chiffre d'affaires. Valeur de Shapley des datasets. Identification des auteurs dans chaque dataset, assortis de la valeur Pay-to-train. Traitement et contrôle par le Bureau de l'IA : Observation des méthodes de calcul (valeurs de Shapley et Pay-to-train) et de recoupement (identification des ayant-droits). Fixation des normes de présentation des données (ex. ISWC). Définition et mode de groupement des ayant-droits représentant peu d'œuvres. Publication du guide des bonnes pratiques : valeurs de répartition, méthodes de calcul, méthodes d'identification des auteurs, coût de l'équité. Pénalités applicables en cas de fausse déclaration	à notre avis, l'URL[128] des sites n'est pas nécessaire si les fournisseurs de modèles calculent et déclarent les valeurs pay-to-train
6 - Collecte de la redevance équitable	Estimation, par le Bureau de l'IA, du chiffre d'affaires des contenus générés, par pays européen Déduction forfaitaire, par le Bureau de l'IA, pour les datasets d'auteurs provenant de la « gig economy » ou des banques de media. Ce forfait devrait être établi selon les valeurs de Shapley de ces datasets et un plafond pour maîtriser le risque d'attrition vers la gig economy. Collecte : par un organisme désigné par le Bureau de l'IA	Spré, ASCAP
7 - Association OGC / dataset	Listage par le Bureau de l'IA de tous les datasets utilisés par tous les fournisseurs de modèles, et diffusion aux OGC Revendication d'attribution des datasets par les OGC Attribution des datasets aux OGC par le Bureau de l'IA	

8 - Liste par auteur	Diffusion, par le Bureau de l'IA aux OGC, des répartitions par auteur en fonction des datasets attribués	Les OGC sont tenues au secret
9 - Répartition par OGC	Répartition de la redevance équitable entre OGC, selon les datasets attribués et au pro rata des valeurs de Shapley, par l'organisme de collecte Accords de coordination ou de réciprocité entre OGC[129]	
10 - Répartition par auteur	Répartition de la redevance équitable entre auteurs, selon les valeurs pay-to-train, par les OGC	à notre avis, il n'est pas nécessaire de calculer les valeurs pay-to-train œuvre par œuvre[130]
11 - Gestion des redevances non attribuées	Mise en réserve, par les OGC, des redevances non attribuées Recherche des ayant-droits légitimes : consultation de bases de données, demande d'informations, portail en ligne, appel à revendication Réaffectation des montants restants	L324-14 du CPI[131], œuvres non identifiées[132], sommes non répartissables[133]
12 - Coût de gestion	Retenue pour coût de gestion, par les fournisseurs de modèles, pour financer la recherche et la mise en œuvre (méthode et calculs de répartition), fixée par le Bureau de l'IA après concertation avec les OGC. Retenue pour coût de gestion, par l'organisme de collecte, définie par le Bureau de l'IA Retenue pour coût de gestion, par les OGC, définie en interne.	

La figure 16 illustre les flux financiers à partir d'une collecte pour rémunération équitable représentant par exemple 15% du chiffre d'affaires par les contenus générés.

Figure 16 – Schéma des flux financiers (chiffres indiqués à titre illustratif)



20 – Mise en perspective

Nos recommandations proposent une solution « bêta » de rémunération équitable des auteurs par l'IA générative.

Cette solution est fondée sur la propriété intellectuelle des auteurs. Elle impose une redevance globale pour les IA, une traçabilité des sources y compris lorsque leur contribution est minime. Elle reconnaît la créativité et la distingue du simple contenu amateur. Elle met en œuvre des mécanismes d'incitation individuelle pour cultiver l'ensemble du capital créatif.

Au-delà de la créativité, cette solution s'applique aux savoir-faire, dans toutes les filières économiques. Elle solvabilise les pools de propriété intellectuelle pour lutter contre la dynamique des monopoles et de l'IA ubérisante. Elle ouvre une perspective et une méthode pour un projet européen de revalorisation du travail et de son capital humain et culturel.

Elle révèle enfin une possible position stratégique de l'Europe dans les nouvelles chaînes de valeur, de manière réaliste au regard de ses faiblesses dans le financement de l'innovation, et compatible avec ses aspirations humanistes et sociales.

21 – Remerciements

Cette note intermédiaire s'inspire des réflexions et échanges entre les membres des missions ministérielles pour l'AI Act, ceux des think tanks indépendants Controv3rse et Institut de l'Iconomie, ainsi que d'autres personnalités reconnues pour leur expertise. Elle prolonge l'étude[134] faite précédemment pour la mission parlementaire Pradal-Rambaud[135] et des tribunes publiées dans *Le Monde* et *Les Echos*. Vincent Lorphelin, Fondateur et co-Président de ces think tanks, rédacteur de cette note, les remercie chaleureusement pour leurs idées et la qualité de ces échanges.

Présidents et Rapporteurs des missions parlementaire et ministérielle :

Alexandra Bensamoun, professeure de droit à l'université Paris Saclay, ancienne membre de la Commission de l'intelligence artificielle

Joëlle Farchy, professeure des universités, Directrice du m2ecn et de la Chaire PcEn Université Paris 1 Panthéon-Sorbonne, membre du CSPLA

Philippe Pradal, Député de la Troisième Circonscription des Alpes Maritimes, Assemblée Nationale

Stéphane Rambaud, Député de la Troisième Circonscription du Var, Assemblée Nationale

Lionel Ferreira, Maître des requêtes au Conseil d'Etat

Julie Groffe-Charrier, Maître de conférences HDR en droit privé chez Faculté Jean-Monnet

Bastien Blain, Professeur Junior en économie, scientifique des données et du comportement, Université Paris 1 Panthéon-Sorbonne

Experts et Entrepreneurs :

William Bailey
co-Fondateur et président, Bolero Music

Emmanuel Benazera
Président Directeur Général, Jolibrain

Jean-Paul Betbèze
Economiste, Betbèze Conseil

Michelle Bergadaà
Présidente à l'Institut de Recherche et d'Action sur la Fraude et le Plagiat Académiques

Michel Bokobza
Chargé de missions, Collège de Paris

Laurent Cervoni
Directeur de la Recherche et de l'Innovation, Talan

Charles-Éric de la Chapelle
Fondateur, Myriad Data

Jean-Philippe Clair
Directeur Marketing, Communication et Innovation, Keyrus

Alice Coatalem
co-Fondatrice de CogNeed et Professeur Associée, Université Paris Dauphine PSL

Christophe Collet
Fondateur, AskLocala

Frédéric Dayan
Fondateur, Exactcure

Guillaume Desveaux
co-Fondateur d'Aleia et Administrateur, AI Cargo Foundation

Jonathan Dory
Fondateur, Live Crew

William Eldin
Fondateur, XXII

Pierre Fernandez
Doctorant en informatique, Inria Rennes

Stéphanie Flacher
co-Fondatrice, Logion Network

Raphaël Frisch
co-Fondateur, HawAI.tech

Alain Garnier
CEO, Jamespot

Anthony Graveline
Fondateur, Disaitek

Philippe Guguen
Président, Sorga - Map Emulsion

Francis Hachem
Fondateur, Codenekt

Rodolphe Hasselvander
Fondateur, Blue Frog Robotics

Brice Hoarau
Fondateur, Semdee

Matthias Houllier
co-Fondateur, Wintics

Casey Joly
Avocate, IPso

Jean Latger
Fondateur, Oktal-SE

Frédéric Lefebvre-Naré
Directeur data IA, Niji

Youness Lemrabet
Fondateur, Everysens

Jacques Lévy-Vehel
Président de Case Law Analytics, et Directeur de Recherches, INRIA

Olivier Laborde
Leader Innovation et Transformation digitale, BPCE

Jean Latger
Président Directeur Général, Oktal-SE

Alexandre Leforestier
Fondateur, Panodyssey

Sixtine Lorphelin
Ingénieure IA, UTC, étudiante à l'INSEAD

Aymeric Masurelle
co-Fondateur, Spoon.ai

Pierre Miralles
co-Fondateur, Footovision

Emmanuel Moyrand
co-Fondateur, France Meta

Clément Merville
Fondateur, Manzalab - Teemew

Edouard de Miollis
Fondateur, Polycube

Nathalie Nevejans
Professeure de droit privé et intelligence artificielle, Université d'Artois

Rémy Ozcan
co-Fondateur, Crypto4All

Jean-Jacques Quisquater
Professeur de Cryptographie, Polytechnique Louvain, chroniqueur du Monde

Jean-Michel Salomon
Président, Société des Auteurs de Jeux

Frédéric Soufflet
co-Fondateur, Haapie

Clément Tequi
co-Fondateur, Ternoa

François-Xavier Thoorens
CEO, Vaultys

Arnaud Touati
Avocat Associé, #Hashtag

Christophe Tricot
co-Fondateur, La Forge

Killian Vermersch
co-Fondateur et Directeur Général,
Golem.ai

Bibliographie