MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

Rémunération des contenus culturels utilisés par les systèmes d'intelligence artificielle

Mai 2025

MISSION CONFIÉE PAR
LE CONSEIL SUPÉRIEUR DE LA PROPRIÉTÉ
LITTÉRAIRE ET ARTISTIQUE À
ALEXANDRA BENSAMOUN
ET JOËLLE FARCHY

Projet de rapport - Volet économique

Présenté le 23 juin 2025 à la séance plénière du CSPLA

Joëlle Farchy, co-présidente

Bastien Blain, co-rapporteur

Rémunération des contenus culturels utilisés par les systèmes d'IA

Rapport économique

En France, la ministre de la Culture a souhaité que le Conseil Supérieur de la Propriété Littéraire et Artistique (CSPLA) prenne en charge une mission « relative à la rémunération des contenus culturels utilisés par les systèmes d'intelligence artificielle ».

Par une lettre en date du 12 avril 2024, la présidence de cette mission a été confiée aux professeures Alexandra Bensamoun pour le volet juridique et Joëlle Farchy pour le volet économique. Par la suite, un questionnaire a été envoyé à des professionnels par les rapporteurs de la mission, Bastien Blain et Julie Groffe. Les réponses à ce questionnaire ont permis d'alimenter les premières réflexions.

Le rapport économique, présenté ici, ne vise pas à apporter des solutions définitives à l'ensemble des questions soulevées. En posant quelques jalons, l'objectif est de nourrir la phase d'approfondissement qui s'amorce dans la perspective de travaux ultérieurs.

Le contenu de ce rapport n'engage que ses auteurs.

Joëlle Farchy

Avec Bastien Blain

Avril 2025

Rapport économique 1
1 - Alimenter le patrimoine culturel de l'humanité pour le préserver de la maladie de « l'œuvre folle » 5
1.1 - De l'inspiration à la substitution : la crainte du grand remplacement 5
1.2 - Dégénérescence des modèles entrainés sur des données synthétiques 8
2 – Transferts de valeur entre opérateurs d'IA et titulaires de DPI : cadre de mise en œuvre 15
2.1 Liberté contractuelle et négociations directes 15
2.1.1 Les négociations dans le cadre de l'exception TDM 15
2.1.2 Les négociations dans le cadre d'initiatives contractuelles indépendantes 16
2.1.3 Les limites de la contractualisation dispersée 16
2.2 Les transferts de valeur obligatoires 17
2.3 L'hypothèse d'une voie complémentaire : accompagner la construction d'une place de marché 19
3 – Chaîne de valeur et acteurs des systèmes d'IA 22
3.1 Typologie des systèmes et modèles 22
3.2 Les segments de la chaine de valeur 23
3.2.1 Les ressources –données 23
3.2.2 Le développement : les étapes de la modélisation 24
3.2.3 Lancement des modèles et déploiement 26
3.2.4 Usagers 27
3.2.5 Des circuits complexes de valorisation dans la phase de déploiement 27
3.3 Opérateurs et tendances du marché 29
3.3.1 L'essor des partenariats 29
3.3.2 Un oligopole d'entreprises américaines domine le marché du développement des modèles de fondation 30
3.3.3 Small is beautiful 31
4 – Valorisation des données-œuvres pour les systèmes d'IA 33
4.1 Quantifier la valeur des données 33
4.1.1 Etablir des liens de causalité en modifiant les paramètres des modèles 33
4.1.2 Etablir des liens de similarité entre la sortie du modèle et les données d'entraînement - Méthode « passive » corrélationnelle 36
4.1.3 Marquage des données d'entrainement - Méthode proactive causale 37
4.2 Rémunération des œuvres protégées par le droit d'auteur : les étapes de la valorisation 38
4.2.1 Une assiette de rémunération liée à l'activité de l'exploitant 38
4.2.2 Une part attribuée à l'amont modulée selon les habitudes professionnelles et les rapports de force entre les acteurs 39

4.2.3 La répartition entre les œuvres et les ayants droit 40
4.3 Quantifier les transferts de valeur dans le cas de l'IA 40
4.3.1 Lignes directrices 40
4.3.2 Apport opérationnel des méthodes de quantification 43
Annexes 51
Encadré 1 - Modèles de langage, modèles de diffusion et mesures de l'effondrement 52
Encadré 2 – Les sources d'erreurs du processus d'effondrement 55
Encadré 3 : La méthode de la « Shapley value » appliquée à des jeux de données culturelles protégées 56
Encadré 4. Les différentes méthodes d'estimation de la contribution de données d'entraînement à la sortie de l'IA générative. 58
Bibliographie 61

Pour effectuer les diverses tâches nécessaires à leur fonctionnement, les systèmes d'IA, ont besoin de multiples données. Parmi elles, peuvent se trouver des données-œuvres protégées par la propriété intellectuelle. La voracité de ces modèles ne saurait cependant être un argument en faveur d'un recours généralisé à une consommation débridée pour l'ensemble des œuvres protégées. Les acteurs culturels expriment en effet la crainte que le non respect des règles de propriété intellectuelle conduise à ce que le financement de la création ne soit plus assuré.

Le droit de la PI a été conçu autour de l'œuvre, objet envisagé dans son unité et sa singularité. Les instruments classiques d'autorisations et de rémunération peuvent souvent paraître inappropriés face aux approches volumétriques de l'IA, en rupture avec les mécanismes d'autorisation conçus au regard d'exploitation individuelle d'objets déterminés alors que les notions de données et de contenus obéissent à des mouvements inverses de flux permanents et de grandes masses de l'économie numérique (Benabou, 2018).

C'est pourquoi **des solutions pragmatiques, adaptées à des périmètres vastes, doivent être trouvées, tout en respectant les principes généraux associés au DPI. L'objectif est de valoriser les données – œuvres européennes au sein d'un écosystème garantissant à la fois leur circulation et la pérennité de leur financement.** Dans un contexte de concurrence internationale intense, l'innovation dans l'IA est en effet devenue un enjeu de compétitivité majeur sur le plan de la souveraineté industrielle mais aussi culturelle. Seule une large mobilisation des idées et des contenus européens permettra, en effet, de limiter le risque culturel majeur que les systèmes d'IA ne proposent plus, in fine, que des contenus dont les références culturelles européennes seraient totalement exclues.

Afin d'éclairer les enjeux, ce rapport est construit en quatre temps. Dans une première partie nous montrons pourquoi l'incitation à l'investissement dans la création humaine, via des transferts de valeur, est nécessaire à la fois pour les acteurs culturels et pour soutenir l'innovation dans l'IA. La seconde partie indique le cadre de mise en œuvre possible des transferts de valeur. Après avoir présenté dans la partie suivante les étapes de la chaîne de valeur et les lieux de création de valeur pour les acteurs des systèmes d'IA, nous abordons dans la dernière partie, la valorisation des données œuvres pour les systèmes d'IA et les lignes directrices de ce que pourrait être le partage de la valeur au profit des opérateurs culturels.

1 - Alimenter le patrimoine culturel de l'humanité pour le préserver de la maladie de « l'œuvre folle »

Dans un marché mondialisé et concurrentiel, l'idée que les titulaires de droits puissent être associés au partage de la valeur des systèmes d'IA n'est pas toujours clairement perçue. C'est pourquoi cette première partie vise **à mettre en évidence les raisons économiques susceptibles de justifier des transferts de valeur entre opérateurs de modèles d'IA et titulaires de DPI.** Les opérateurs ont en effet besoin de recourir à certaines données protégées comme le soulignent des représentants d'Open AI dans le cadre d'une enquête de la commission *Communications and Digital* de la *House of Lords* au Royaume-Uni (House of Lords Communications and Digital Select Committee inquiry, 2023) : *« We believe that AI tools are at their best when they incorporate and represent the full diversity and breadth of human intelligence and experience. In order to do this, today's AI technologies require a large amount of training data and computation, as models review, analyze, and learn patterns and concepts that emerge from trillions of words and images. OpenAI's large language models, including the models that power ChatGPT, are developed using three primary sources of training data: (1) information that is publicly available on the internet, (2) information that we license from third parties, and (3) information that our users or our human trainers provide. Because copyright today covers virtually every sort of human expression – including blog posts, photographs, forum posts, scraps of software code, and government documents – **it would be impossible to train today's leading AI models without using copyrighted materials.** Limiting training data to public domain books and drawings created more than a century ago might yield an interesting experiment, but would not provide AI systems that meet the needs of today's citizens ».*

1.1 - De l'inspiration à la substitution : la crainte du grand remplacement

Il serait possible d'envisager que compte tenu de la masse de données entrantes mobilisées par les modèles, plus que des classiques questions d'exploitation d'une œuvre, nous serions en présence d'un acte « d'inspiration », comme un humain lit des centaines de livres ou écoute des milliers de morceaux de musique pour écrire ou composer lui-même.

Cependant, pour la plupart des acteurs culturels, l'utilisation d'œuvres existantes pour entraîner une intelligence artificielle n'est pas comparable aux réminiscences qui alimentent la création humaine. Si un auteur s'inspire souvent d'œuvres créées par d'autres, sans que cela puisse toujours donner prise au droit d'auteur, la systémisation et l'ampleur tant quantitative que qualitative que constitue l'intelligence artificielle modifient totalement la portée du phénomène. Les questions juridiques de respect du monopole d'exploitation et d'accès licite aux œuvres sont traitées par ailleurs dans le rapport. La question économique à laquelle nous nous attachons dans cette partie est de savoir si un acte réalisé grâce à des œuvres entrantes pour entraîner ou améliorer la performance d'un modèle d'IA, doit être rémunéré sur la base d'une éventuelle perte de valeur subie.

Dans le cas des actions effectuées par les systèmes d'IA, l**a vraie disruption tient**, non pas seulement au changement d'échelle des actes effectués, ou à l'accès massif ou à l'exploitation technique des œuvres mais au risque de « grand remplacement » c'est-à-dire **au fait que de nombreux résultats générés par IA s'apparentent à ce que l'on peut nommer des « quasi-œuvres »** (Benabou, 2023) **et concurrencent ainsi directement les créations humaines ayant servi à leur élaboration** (**cf. Figure 1, point 1) ce qui, à terme, pourrait compromettre les conditions mêmes de la création humaine.**

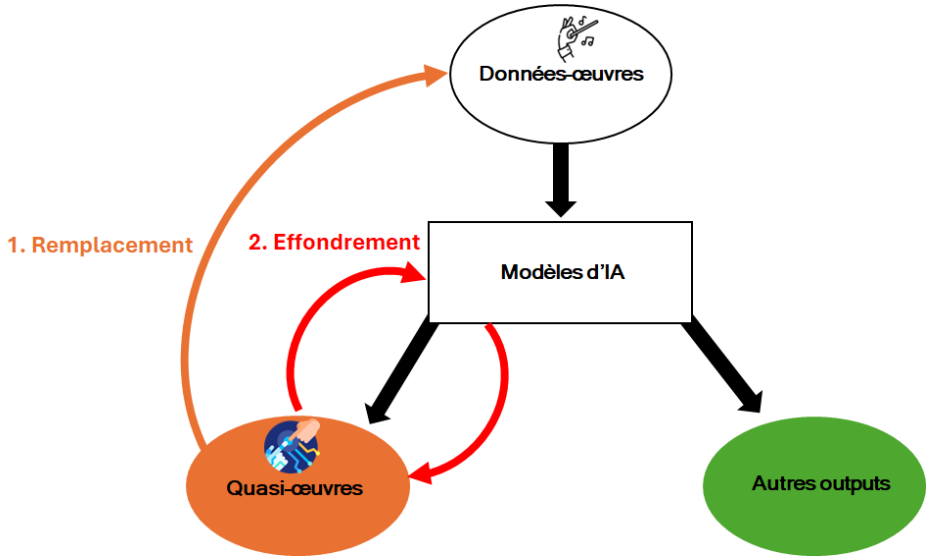


Figure 1. Circuit d'alimentation des modèles d'IA par des données œuvres

La crainte du grand remplacement des humains par des machines pose donc avec une acuité particulière la question des transferts de valeur. Soulignons que la perte de valeur subie en cas de mobilisation des données œuvres pour l'IA ne s'analyse pas seulement comme celle, classique et directe, d'un auteur particulier qui est « copié » lorsque l'output produit ressemble de trop près à l'input ingéré ou celle d'un acteur économique ayant investi dans la création. La perte repose également ici, sur le degré de substituabilité entre œuvres humaines et synthétiques au niveau macro-économique. Elle correspond, de manière tangible, aux effets de la concurrence sur des ensembles de mondes professionnels et sur l'avenir de ces mondes.

Les effets d'éviction sur la création humaine s'exercent d'abord par une concurrence par les prix, en ce que l'IA permet de créer des *outputs* plus vite et de manière moins coûteuse que des humains. C'est par exemple le cas pour les traductions, passées d'environ 21 euros le feuillet de 1500 signes à 17 ou 18 euros (Vulser, 2024). Les effets d'éviction s'exercent également par les quantités. La surabondance d'une offre générée par l'utilisation des systèmes d'IA risque en effet d'impliquer une saturation du marché et par conséquent, une moindre découvrabilité des œuvres humaines par l'usager. C'est par exemple le cas de la prolifération des clones vocaux non autorisés, en particulier sur YouTube et TikTok, qui cause un problème économique en détournant l'attention des enregistrements officiels.

Les effets de l'IA sur l'emploi ont donné lieu à diverses études. Les effets semblent limités à court terme mais 33% des emplois des économies avancées pourraient être remplacés par l'IA à long terme, d'après un rapport du FMI (Cazzaniga et al., 2024). En mars 2023, des chercheurs d'OpenAI, d'OpenResearch et de l'Université de Pennsylvanie ont estimé que les grands modèles de langage (*LLM*, *Large Language Model*) pourraient affecter les fonctions professionnelles de 80 % de la main-d'œuvre aux États-Unis (Eloundou et al., 2023). De plus, des chercheurs du Massachussetts Institute of Technology, de la London School of Economics et de l'Université de Boston ont détecté une corrélation négative entre l'adoption de l'IA et le recrutement d'emplois entre 2010 et 2018 : pour chaque augmentation de 1 % de l'utilisation de l'IA, les entreprises exposées à l'IA ont réduit leurs embauches d'environ 1 % (Acemoglu et al., 2022). Selon d'autres estimations, effectuées par la Direction Générale du Trésor, portant plus précisément sur l'arrivée des modèles de fondation, si 80 % des travailleurs américains pourraient voir au moins 10 % de leurs tâches remplacées,

seulement 19 % d'entre eux pourraient voir cette part atteindre au moins 50 %, et feraient donc face à un risque important de substitution (Besson et al., 2024).

L'adoption de l'IA menacerait davantage les professions les plus qualifiées (diplômés du supérieur avec des salaires élevés), en se substituant à certains travailleurs hautement qualifiés pour réaliser des tâches qui requièrent des compétences avancées. En effet, l'IA est capable de prendre en charge des tâches cognitives abstraites et non-routinières, et donc d'élargir le périmètre des tâches substituables (e.g. traduction, élaboration de diagnostics). Parmi les professions qualifiées, toutes ne devraient pas être touchées dans les mêmes proportions. Les entreprises pourraient davantage diminuer leurs effectifs dans les professions axées sur l'écriture et la programmation, plus exposées au risque de remplacement par les modèles génératifs.

Dans les secteurs culturels, certains métiers se sentent d'ores et déjà particulièrement menacés. Le métier de **traducteur** sous sa forme actuelle est remis en cause. Les traducteurs reçoivent de moins en moins de demandes de traductions complètes et davantage de travaux de prestations consistant à corriger une traduction produite par un système d'IA tel que DeepL. Ce travail de post édition est parfois jugé plus chronophage et conduit à une rémunération moindre, selon une enquête sur la traduction automatique et la post-édition, menée par l'Association des traducteurs littéraires de France (ATLF), en décembre 2022, auprès de quatre cents personnes (*L'ATLF a Interrogé Ses Adhérents Sur La Post-Édition*, 2022). En effet, la rémunération est moindre et pour 68 %, elle était même inférieure aux tarifs moyens de traduction (Vulser, 2024). Si la littérature, qui représente moins de 10% de la production éditoriale en France chaque année est plutôt épargnée, les ouvrages laissant moins de place à l'interprétation sont concernés. Au contraire, dans le monde de la BD ou des *webtoons* (BD en format smartphone), des outils d'IA comme Geo Comix sont utilisés pour traduire les bulles dans plusieurs langues. Il en va de même pour les livres audio, avec l'utilisation de système d'IA par Harper Collins, tandis qu'Audible propose de nombreux livres dont la voix est générée par une IA (Cohen, 2024). Par ailleurs, de nombreuses page web sont traduites avec des IA (Thomson et al., 2024).

Les **graphiste**s sont fortement menacés par des systèmes tels Midjourney, avec la possibilité de générer par exemple des illustrations de science-fiction. Ainsi, la *Society of Authors* au Royaume-Unis a estimé via un sondage qu'environ un quart des illustrateurs ont déjà perdu du travail en raison de l'IA générative et plus d'un tiers des illustrateurs (37 %) déclarent que leurs revenus ont diminué en valeur à cause de l'IA générative (*SoA Survey Reveals a Third of Translators and Quarter of Illustrators Losing Work to AI - The Society of Authors*, 2024)

En ce qui concerne les **métiers de l'écriture**, comme le **journalisme**, une étude récente estime que les métiers de reporters et journalistes sont les plus exposés aux systèmes d'IA (Eloundou et al., 2023). Plusieurs médias, tels que BuzzFeed, News Corp Australia et G/O Media, ont intégré l'IA générative dans leur production de contenus. Début 2023, BuzzFeed a lancé des quiz alimentés par ChatGPT, des articles de voyage et un *chatbot* de recommandation de recettes nommé Botatouille (Chin-Rothmann, 2023). Parallèlement, Google a proposé à des médias nationaux et locaux un outil nommé Genesis, *chatbot* génératif capable de rédiger des titres, des publications sur les réseaux sociaux et des articles, présenté comme un outil visant à améliorer la productivité. C'est ainsi que plus d'une centaine de sites d'informations et d'actualités entièrement créés par des IA ou presque ont été identifiés depuis 2023 (Sadeghi et al., 2024). Par ailleurs, de nombreux livres sont écrits via des système d'IA, comme témoigne le nombre important d'*ebooks* autopubliés sur des plateformes telles que Kindle, dont le nombre est désormais limité à trois par jour et par auteur (*Update on KDP Title Creation Limits*, 2023).

Pour les doubleurs, une étude du Datalab du groupe Audiens (*Audiens data lab report*), quantifie les métiers menacés en France par les outils de doublage fournis par des systèmes d'intelligence artificielle. En 2023, cette industrie concerne 110 entreprises et a employé 7397 intermittents du spectacle et 3116 permanents. Les systèmes d'IA comme HeyGen, Eleven Dubs ou Deepdub,

permettent de cloner des voix et de traduire des vidéos en plusieurs langues tout en adaptant les mouvements de lèvres. Leur utilisation permettrait d'éviter l'enregistrement en studio des doublages de films, séries, jeux vidéo et dessins animés par des comédiens. Cette délocalisation et cette automatisation pourraient engendrer une perte d'activité massive dans ce secteur (Thomas, 2023). Cette inquiétude ne concerne pas que les doubleurs français, dans la mesure où les doubleurs de jeux-vidéos se sont mobilisés, par exemple en Californie. En effet, 2600 artistes qui assurent le doublage de jeux vidéo, ou dont les mouvements servent à animer les personnages de synthèse pourrait voir les systèmes d'IA reproduire la voix d'un comédien ou créer une réplique numérique d'un cascadeur, sans son consentement ou sans rémunération (*Inquiets de l'utilisation de l'IA, les acteurs et doubleurs de jeux vidéo vont faire grève en Californie*, 2024).

Dans le cas du doublage, la question de la concurrence entre doubleurs humains et doubleurs IA, qui alimente les réactions de la profession, se conjugue avec une préoccupation de politique industrielle nationale si le doublage via l'IA devait se généraliser. En effet, depuis l'émergence du cinéma parlant, Hollywood a affiché pour ambition de capter le marché du doublage sur son sol, ce qu'elle n'est jamais parvenue à faire, une véritable industrie performante du doublage s'étant développée en France. Or, dans la mesure où les acteurs de l'IA qui s'imposent sur le marché du doublage (Open Ai, Eleven Labs…) sont américains, le doublage par IA a toutes les chances de se faire au profit des seuls studios américains, sur le sol américain.

Outre ce débat sur la souveraineté industrielle, celui sur l'articulation entre menaces à court terme sur l'emploi et adaptation à plus ou moins long terme entre apports de l'humain et de la machine est loin d'être clos[1]. Il renvoie, dans une perspective à la Schumpeter, au processus macro-économique de destruction créatrice associé à toute innovation. A court terme, durant une nécessaire période de transition, **des politiques d'accompagnement sur le plan social, qui tiennent compte de l'expertise et des savoirs faire accumulés, seront indispensables.** De plus, dans la mesure où de nombreux métiers sont amenés non à disparaître mais à se transformer profondément, la **question de la formation** devient essentielle afin que les professionnels de demain ne subissent pas cette innovation mais en prennent le contrôle. C'est pourquoi la mission alerte les Pouvoirs publics sur l'urgence de la mise en place ou de la consolidation de dispositifs adaptés pour les activités directement concernées.

1.2 - Dégénérescence des modèles entrainés sur des données synthétiques

Au-delà des effets à court terme sur la filière culturelle, le **risque de remplacement de la création humaine par l'IA pourrait, de plus, conduire, à moyen ou long terme, à une contradiction interne aux modèles d'IA eux-mêmes, et à leur possible effondrement.**

Une fois qu'un modèle est entraîné sur des données humaines, réelles, il est possible de l'utiliser pour générer de nouvelles données, appelées synthétiques. Les données synthétiques sont conçues pour imiter les propriétés statistiques et structurelles des données réelles, tout en étant générées artificiellement. Dans le cas des IA génératives actuelles, il s'agit des données créées par le modèle entraîné initialement sur des données humaines, sur demande, comme des images de paysages. De ce fait, on peut, en principe, entraîner un modèle uniquement sur des données synthétiques, ou en les combinant avec des données humaines. Autrement dit, les données de sortie (i.e. l'output) d'un modèle entraîné sur des données créées par des humains peut être utilisé comme données d'entrée (i.e. input) pour un nouveau modèle. Néanmoins, procéder ainsi pourrait

[1] Dans une tribune d'un collectif publié dans le Monde daté du 10 septembre 2024, une citation du traductologue américain Alan Melby indique que l'IA ne devrait pas faire disparaître le métier de traducteur « sauf peut-être pour ceux qui traduisent d'ores et déjà comme des machines »(*« Non, l'intelligence artificielle ne remplacera pas les traducteurs et les traductrices ! »*, 2024).

conduire à générer des données de mauvaise qualité, c'est-à-dire à un effondrement du modèle (cf. *Figure 1, point 2).*

L'impact de l'entraînement des modèles d'IA via des données synthétiques (plutôt que produites par des humains) sur la qualité des données générées par ces modèles, a fait l'objet de divers travaux. Ces derniers s'appuient sur des mesures de la qualité des résultats (cf. Encadré 1 en Annexes) et identifient les sources d'erreurs qui peuvent être à l'origine de l'effondrement (cf. Encadré 2 en Annexes).

Dans une étude récente publiée dans la prestigieuse revue *Nature*, des chercheurs d'Oxford et Cambridge (Shumailov et al., 2024) ont démontré l'effondrement de la qualité des *outputs* d'un modèle affiné sur des données synthétiques. Le terme d'effondrement est ici défini comme un process dégénératif affectant la qualité de ce que produisent les modèles, dans lesquels les données générées par une première génération de modèles polluent les données sur lesquelles la prochaine génération de modèles est entraînée. Dans cette étude, les chercheurs ont affiné un modèle de langage (un *LLM*) pré-entraîné sur un ensemble de données émanant de Wikipédia (données créées par des humains), qu'ils ont ensuite utilisées pour générer de nouveaux articles (données synthétiques). Ils ont formé la génération suivante du modèle sur la base de ces nouveaux articles plutôt que sur des données réelles, et ainsi de suite. Lors de l'évaluation, les nouveaux modèles ont rapidement montré des erreurs significatives par rapport au modèle original entraîné sur de réelles données. En d'autres termes, la qualité des nouvelles données s'est effondrée après que les modèles aient été entraînés sur des données synthétiques.

L'effondrement se produit parce que chaque modèle ne s'appuie que sur des données synthétiques, ce qui conduit à un renforcement des mots communs et à une négligence des mots rares. Au fil des itérations, le modèle apprend de plus en plus de ses propres prédictions erronées, amplifiant les erreurs jusqu'à ce qu'il apprenne essentiellement sur la base d'informations incorrectes. L'effondrement correspond à une perte progressive des informations sur la distribution réelle des données ; les modèles deviennent de moins en moins capables de produire des sorties diversifiées, avec une réduction de la variance de leurs distributions de sortie. Les événements rares ou peu probables disparaissent des connaissances du modèle à mesure qu'il continue de s'entraîner sur ses propres données. Dans le même temps, au fil des générations, des données très improbables sont générées, données que le modèle de première génération n'aurait jamais produites : des données aberrantes. L'effondrement se traduit par des répétitions plus importantes et l'introduction d'erreurs. Le modèle finit par mal percevoir la distribution sous-jacente des données, se concentrant sur ses propres projections de plus en plus inexactes du monde et générant des données aberrantes.

Une étude de chercheurs de l'université de Standford et de Rice confirme que le même phénomène survient dans le cadre de la génération d'images (Alemohammad et al., 2023) qui repose sur des modèles de diffusion (l'« équivalent » des grands modèles de langage pour les images). Dans cette étude, les chercheurs étudient l'impact de boucles d'entraînement de modèles de génération d'images, différentes dans leur manière de combiner données réelles et données synthétiques, et montrent que la qualité et la diversité s'effondrent lorsque seules des données synthétiques sont utilisées.

Une étude similaire menée par des chercheurs de Standford et Berkeley (Bohacek & Farid, 2023) confirme ce résultat pour un autre type de modèle permettant de générer des images, le populaire Stable Diffusion (Rombach et al., 2022). Bohacek & Farid, prenant l'exemple de la génération de visages, montrent que les images produites par le modèle de base sont d'excellente qualité ; mais l'effondrement de leur qualité est observé dès que le pourcentage de données synthétiques utilisées pour l'entraînement excède 10%.

Le risque d'effondrement est à prendre au sérieux dans la mesure où les données sont souvent extraites d'Internet et que celui-ci regorge de plus en plus de données synthétiques (Alemohammad et al., 2023), comme des images, des évaluations (Gault, 2023), des sites web (Cantor, 2023) ou des données annotées (Veselovsky et al., 2023). Certaines bases de données populaires d'images contiennent des données synthétiques dont l'usage est parfois volontaire, parfois lié au manque de données accessibles comme en médecine (Pinaya et al., 2022) ou en géophysique (C. Deng et al., 2022) ou du fait de la protection des données médicales privées (Klemp et al., 2023; Luzi et al., 2024).

De plus, le recours aux données synthétiques pourrait devenir une conséquence de la raréfaction de données crées par des humains (Villalobos, 2022). En supposant que les taux actuels de consommation et de production de données se maintiennent, **les données réelles vont manquer.** En effet, des recherches menées par Epoch AI prédisent que « nous aurons épuisé le stock de données textuelles de faible qualité d'ici 2030 à 2050, les données textuelles de haute qualité avant 2026, et les données visuelles entre 2030 et 2060. » (Villalobos, 2022). Il n'est par ailleurs pas vraiment possible d'entraîner davantage les modèles sur les données existantes à cause du risque « d'*overfit* » (i.e. entrainer le modèle à expliquer des variations stochastiques propres au jeu de données d'entrainement et à lui seul), c'est-à-dire du risque d'altérer la généralisation du modèle et donc sa capacité à générer de nouvelles données.

Face à ces risques avérés, comment prévenir l'effondrement des modèles ? Une étude montre que, dans le cadre du remplacement des données réelles par les données synthétiques, l'effondrement ne se produira pas si les modèles génératifs initiaux approximent suffisamment bien la distribution des données réelles et si la proportion de données réelles est suffisamment grande par rapport aux données synthétiques (Bertrand et al., 2024). De manière similaire, (Dohmatob et al., 2024) suggèrent que le choix fin de la qualité des données réelles associées avec des données synthétiques peut éviter l'effondrement du modèle. D'autres chercheurs (Alemohammad et al., 2023) suggèrent toutefois que si sélectionner des images de bonne qualité dans les données synthétiques avant chaque entraînement permet d'éviter la dégradation en terme de qualité des données générées par le modèle, cela conduit tout de même à une réduction de la diversité des données générées (tandis que ne pas sélectionner conduit à un effondrement sur ces deux aspects).

Si les investigations sur l'effondrement des modèles comportent toujours l'entraînement de la génération initiale sur des données réelles, elles divergent sur l'approche adoptée pour entrainer les générations suivantes.

L'approche la plus extrême consiste en un « remplacement » total des données réelles par les données synthétiques dans l'entraînement des générations suivantes, qui conduit, dans toutes les études, à un effondrement.

Pour les autres approches, les études diffèrent sur la proportion de données réelles et de données synthétiques dans l'entraînement des prochaines générations (proportion fixe de données synthétiques versus proportion croissante de données synthétiques). De nombreuses études (Bohacek & Farid, 2023; Martínez et al., 2023; Shumailov et al., 2024) montrent que la présence d'une proportion fixe, parfois même faible de données synthétiques conduit à un effondrement des modèles. Dans ces études, la première génération de modèles est entraînée sur les données réelles tandis que les générations suivantes sont entraînées sur des données comprenant une proportion de données synthétiques associées aux données réelles, qui reste constante à chaque génération. Avec cette approche, la seule manière d'éviter l'effondrement pourrait être d'utiliser de nouvelles données réelles sur lesquelles les modèles n'ont jamais été entraînés (Alemohammad et al., 2023).

Une autre manière de combiner données réelles et données synthétiques consiste en l'accumulation de données synthétiques de chaque nouvelle génération de modèles aux côtés d'une proportion fixe de données réelles pour entraîner la nouvelle génération (approche par « accumulation »). Autrement dit, les données synthétiques s'accumulent dans le temps aux côtés des données réelles pour entraîner chaque prochaine génération. Une étude montre que l'effondrement est retardé et s'accompagne d'une réduction de la diversité des données (Alemohammad et al., 2023)(cf. Annexe technique 2). Une autre équipe de chercheurs de l'université de Standford et du MIT (Gerstgrasser et al., 2024; Kazdan et al., 2024) suggère que lorsque les données synthétiques s'accumulent aux côtés des données réelles au lieu de les remplacer, l'effondrement catastrophique est peu probable, à tout le moins après quelques générations. La dégradation de la qualité de ce qui est produit serait beaucoup plus lente et surviendrait seulement en cas de forte disproportion entre (trop peu de) données réelles et (trop de) données synthétiques, ce qui surviendrait dans le cas d'une création trop faible de nouvelles données.

La proportion de données réelles nécessaires varie d'une étude à l'autre, notamment à propos du terme de leur effet. Shumailov et al., (cf. supra) soulignent que dans le cas de leur étude, il faudrait incorporer 10 % de données réelles pour que l'effondrement se produise plus lentement, ce qui constitue une masse importante de données quand on pense que les modèles sont entraînés sur des trillions de données. D'autres études suggèrent que 10% de données synthétiques est suffisant pour conduire à un effondrement (Bohacek & Farid, 2023).

Le risque d'effondrement du modèle reste donc important lorsque la quantité de données réelles devient insuffisante. L'effondrement potentiel ne signifie pas que les grands modèles de langage ou les autres systèmes d'IA cesseront de fonctionner, mais que cela augmentera les coûts de leur développement. À mesure que les données synthétiques se multiplient en ligne, les lois d'échelle qui suggèrent que les modèles s'améliorent avec plus de données peuvent cesser d'être pertinentes, car ces données synthétiques manquent de la richesse du contenu généré à partir de données « réelles », humaines (Wenger, 2024)(cf. Figure 2 pour illustration).

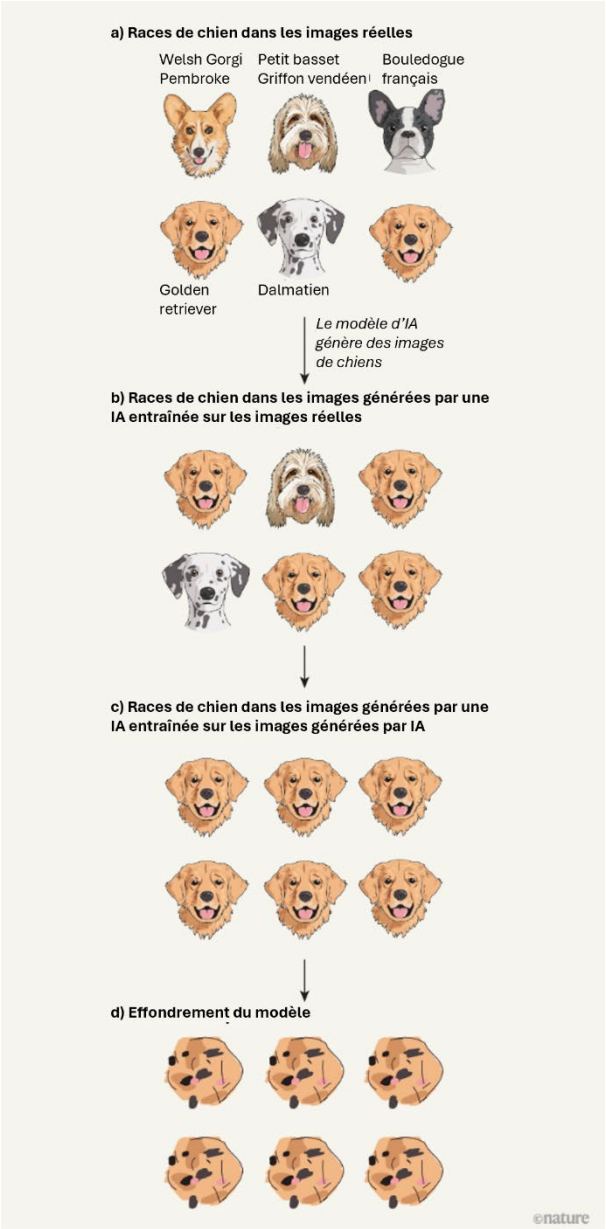


Figure 2. Illustration de la dégénérescence des modèles (adapté de Wenger, 2024)

Le fait qu'entrainer des modèles sur des données synthétiques conduise à une réduction de la diversité des résultats (Alemohammad et al., 2023; Shumailov et al., 2024) est d'autant plus préoccupant dans les cas où les données réelles sont elles-mêmes biaisées et manquent de diversité (Bolukbasi et al., 2016; Caliskan et al., 2017). Certains groupes sont plus représentés que d'autres dans les données (Glickman & Sharot, 2024). En conséquence, l'entraînement sur des données biaisées conduit les modèles à générer des contenus biaisés (Glickman & Sharot, 2024). Un exemple connu est la difficulté des systèmes d'IA de détection de visages à reconnaitre ceux de personnes non-blanches, car ces modèles sont entraînés sur des corpus peu représentatifs du monde réel (Buolamwini & Gebru, 2018; Geirhos et al., 2022). C'est également ainsi qu'un algorithme d'Apple tendait à assigner plus de crédits à des hommes qu'à des femmes (Nasiripour & Natarajan, 2019), qu'un algorithme de recrutement utilisé pour Amazon valorisait les candidatures masculines plus que féminines (Dastin, 2022), et que le moteur de recherche de Google proposait plus d'images d'hommes que de femmes en réponse à une requête neutre comme « personne », en particulier dans les pays à fortes inégalités de genre (Vlasceanu & Amodio, 2022).

De même, Stable Diffusion, utilisé par des millions d'utilisateurs, générait essentiellement des photos d'hommes blancs quand on lui demandait de fournir des images de professions à hauts revenus, comme médecin ou avocat (Bianchi et al., 2023).

La présence de biais dans les données réelles, amplifié par le recours aux données synthétiques peut conduire à renforcer les biais des utilisateurs eux-mêmes lorsqu'ils interagissent avec des modèles d'IA et fausser leurs prises de décisions ((Skjuve, 2023; Troyanskaya et al., 2020). Par exemple, des individus assistés par un système d'IA biaisé pour effectuer un diagnostic médical reproduisent le biais des modèles dans leur décisions, même une fois que les décisions se font sans assistance (Vicente & Matute, 2023). La tendance de l'humain à être biaisé par des systèmes d'IA eux-mêmes biaisés à l'issue de leur utilisation s'étend à d'autres domaines comme la reconnaissance des émotions sur les visages (Glickman & Sharot, 2024).

De nouvelles données humaines et diversifiées restent donc indispensables pour lutter contre la dégénérescence des modèles.

Parmi ces données humaines, certaines concernent plus spécifiquement celles issues des secteurs culturels et protégées par la propriété intellectuelle. Dans ces secteurs, par analogie à la maladie de la « vache folle », on peut évoquer la maladie des « quasi-œuvres folles ». La maladie de la vache folle renvoie à la contamination via la consommation par les bovins de farines animales, qui a pris de l'ampleur en raison du recyclage des carcasses de bovins malades dans des farines animales données en alimentation à d'autres bovins. L**'IA, en remplaçant la création culturelle humaine, pourrait conduire à ne créer que des « quasi-œuvres folles », synthétiques, qui finissent toutes par se ressembler ce qui, par nature, est étranger aux processus de disruption qui jalonnent toute l'histoire de l'activité artistique.** De plus, ces œuvres humaines doivent elles – mêmes être diversifiées si l'on souhaite éviter la dégénérescence des modèles. L'accès à des données de qualité dans le cadre d'une infrastructure technique adaptée, et reflétant la diversité du monde réel, la diversité des langues, des cultures et des régions du globe apparait donc nécessaire dans l'intérêt de l'ensemble des parties.

Pour que ce patrimoine continue à être alimenté, les investissements dans la production et la création humaines doivent être protégés ; il faut tenir compte des investissements réalisés par les ayants droit pour produire des contenus originaux. Au-delà de la perte de revenus à court terme, le risque à plus long terme est celui d'une absence d'investissements pour faire exister les industries culturelles. En l'absence de financement, l'incitation à la création et à la production de nouvelles œuvres humaines de qualité et diversifiées pourrait se tarir.

Dans le domaine de la photo, une étude sur les contributeurs d'Unsplash, plateforme populaire de photos et d'illustrations libres de droits, qui compte environ 6 millions de contenus de haute qualité (Peukert et al., 2024) met ainsi en lumière l'appauvrissement dans ce cas précis. À l'été 2020, Unsplash a lancé un programme de recherche en intelligence artificielle en publiant un ensemble de données comprenant 25.000 images à usage commercial. L'objectif était d'analyser les réactions des contributeurs, en comparant ceux dont les œuvres faisaient partie de ce jeu de données à ceux dont les œuvres n'étaient pas incluses. Les résultats de l'étude montrent que les contributeurs dont les œuvres ont été utilisées dans ce programme ont quitté la plateforme à un taux plus élevé que d'habitude et ont considérablement ralenti leur rythme de téléversement. Cette tendance est plus marquée chez les photographes professionnels prospères que chez les amateurs. De plus, les utilisateurs affectés ont diminué la variété et la nouveauté de leurs contributions à la plateforme, ce qui pourrait avoir des implications à long terme sur le stock d'œuvres disponibles pour le fonctionnement des systèmes d'IA.

***

Ainsi l'IA en remplaçant les œuvres humaines par des « quasi-œuvres » porte le risque, à court terme de déstabiliser un ensemble de mondes professionnels. A plus long terme, ce grand remplacement pourrait conduire à la dégénérescence des modèles s'ils ne sont plus (ou peu) alimentés par des créations humaines nouvelles. La myopie des acteurs économiques et une vision court - termiste des marchés pourraient conduire à ne pas prendre pleinement conscience des enjeux. L'importance de la culture pour nos sociétés n'a plus besoin d'être démontrée. Et au-delà de toute autre motivation, investir dans la création humaine, issue d'horizons variés, est une nécessité pour les modèles d'IA eux-mêmes.

2 – Transferts de valeur entre opérateurs d'IA et titulaires de DPI : cadre de mise en œuvre

2.1 Liberté contractuelle et négociations directes

2.1.1 Les négociations dans le cadre de l'exception TDM

En Europe, la directive de 2019 DAMUN (*Directive 2019/790 sur le Droit d'Auteur et les droit voisins dans le Marché Unique Numérique*) crée une nouvelle exception au droit d'auteur pour la fouille de textes et de données (TDM, *Text and Data Mining*). La réponse proposée par la directive, pour concilier fouille massive de données et autorisations expresses des ayants droit, prévues par le droit de la propriété intellectuelle, est celle d'une exception au monopole, construite en deux temps.

Dans un premier temps, l'article 3 impose une exception au bénéfice des organismes de recherche et des institutions du patrimoine culturel afin de procéder, à des fins de recherche, à des fouilles dans des ensembles d'œuvres ou d'objets protégés. L'exception est prévue sans mécanisme de rémunération compensatoire. Des ayants droit font valoir la non-conformité avec les conditions posées par la directive, que représenteraient les cas de transferts, par des entités publiques, des résultats de leurs recherches à des fins commerciales.

Dans un second temps, afin d'encourager les usages de la fouille de données, l'article 4 prévoit une autre exception, plus large, pour les usages y compris commerciaux, dans des conditions d'accès licite aux données, sans mécanisme de rémunération compensatoire là non plus. Cet article aboutit à un compromis inédit et fort singulier. Il existe en effet une possibilité de sortie, *d'opt-out,* à cette exception pour les titulaires de droits, ce qui suppose que ceux-ci manifestent explicitement, le cas échéant, leur refus de fouille. Autrement dit, la directive prévoit une exception, puis la possibilité de déroger à cette exception et de retourner au monopole du droit exclusif …

Des interrogations sont vite apparues sur la faisabilité pratique du mécanisme *d'opt-out* prévu (pour une première analyse juridique, voir (Bensamoun & Farchy, 2020). **De plus, cet article a été inséré en fin de parcours législatif, à un moment où rien ne laissait imaginer le développement fulgurant de l'IA générative** ; la première version applicative du robot conversationnel d'Open AI, ChatGPT, n'est intervenue qu'en avril 2022. **La question de savoir si l'exception TDM s'applique aux IA génératives** et des implications économiques et sociétales que cet article aurait dans ce contexte **reste encore largement controversée.**

De nombreux ayants droit ont fait valoir leur faculté d'*opt-out,* pour des motivations parfois différentes ; certains ne souhaitent pas voir leurs œuvres utilisées par des IA ; pour d'autres, cette faculté devrait constituer un outil d'incitation à l'ouverture de négociations. Cependant, les conditions d'application de ce processus complexe font l'objet d'interprétations contradictoires ; de plus, les négociations conduisant à rémunération après exercice du droit *d'opt out* prévu par la directive se révèlent souvent infructueuses. Tout d'abord, parce-que l'exception TDM n'ouvre un espace potentiel de négociation que si les titulaires de droits ont exprimé leur intention d'opt out par des "moyens lisibles par la machine" c'est à dire automatisables ce qui conduit à de réelles difficultés à la fois techniques et d'interprétation de cette disposition. De plus, si l'opérateur enlève les œuvres concernées de son jeu de données, la négociation n'a plus lieu d'être (voir schéma de la Figure 3).

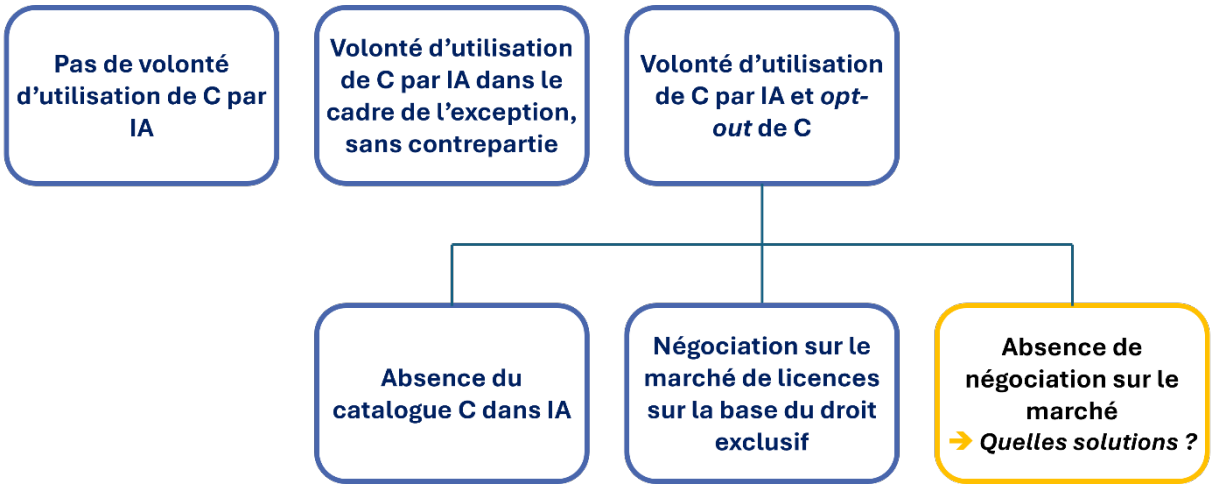


Figure 3. Rémunération d'un catalogue de contenus C par un modèle d'IA dans le cadre de l'exception TDM (en bleu).

A l'heure actuelle, en France, lorsque des réponses aux demandes sont adressées aux ayants droit, ces derniers se voient, très généralement, opposer un refus de négocier ; les arguments invoqués sont divers : les œuvres incriminées ne sont pas utilisées (sans qu'aucune preuve ne soit apportée), la possibilité est offerte de retirer des œuvres si une liste est fournie mais ce retrait ne sera pas rétroactif sur un modèle déjà entraîné, ou encore le fournisseur d'IA a entraîné son modèle dans un pays hors UE, non concerné par la directive. La mise en œuvre de l''exception TDM permet donc, en pratique, difficilement la conclusion de licences et de rémunérations.

2.1.2 Les négociations dans le cadre d'initiatives contractuelles indépendantes

Parallèlement, des entreprises d'IA obtiennent des licences commerciales auprès d'éditeurs de presse afin d'obtenir un accès légal à des données de haute qualité (*Cf.* les accords d'OpenAI avec NewsCorp, aux Etats – Unis, Prisa en Espagne, Axel Springer en Allemagne, Le Monde en France, ou encore Vox Media, Associated Press le Financial Times, les accords de Google avec Reddit, *etc.*, ou dans un autre domaine Adobe et Getty Images). Un accord conclu entre Mistral et l'AFP en janvier 2025 précise que « les contenus AFP ne serviront pas à entrainer et faire progresser les modèles de Mistral. Ces contenus sont un « module » qui vient se brancher au système et peut se débrancher à expiration du contrat ». La plupart des accords ne concernent donc pas les phases de pré- entraînement mais des données d'actualisation, d'ancrage (cf. partie3).

Ces accords, qui prennent des formes multiples (A. Thomas, 2025) , ont en commun d'être établis, selon des initiatives indépendantes visant à fournir des sets de données de qualité et, au-delà, à répondre aux attentes variées et spécifiques de chacun des acteurs.

Des producteurs et éditeurs, notamment dans les secteurs de la presse ou la musique, mettent en avant, pour l'avenir, les avantages d'un marché, libre et compétitif, de licences individuelles volontaires. Ces acteurs soulignent être en mesure d'accorder des licences à grande échelle, comme ils en ont fait la preuve dans le cadre du développement du marché numérique. La libre négociation entre les parties permet notamment aux plateformes de streaming musical de donner accès à la quasi intégralité des répertoires mondiaux.

2.1.3 Les limites de la contractualisation dispersée

Certains risques propres à la conclusion d'accords dispersés entre les entreprises de la tech et des acteurs individuels doivent cependant être soulignés.

Pour les ayants droit, notamment dans certains secteurs, des acteurs sont dans une situation économique qui ne leur permet pas de contractualiser ou de refuser les conditions de rémunérations proposées lors d'accords bilatéraux. Seuls pourront contractualiser ceux dont les données sont les plus convoitées ou ceux qui bénéficieront de l'avantage du « first mover ». Ainsi, une fois qu'Open AI a négocié avec un éditeur de presse renommé dans chacun des pays européens, on voit mal ce qui l'inciterait à négocier avec d'autres acteurs de presse qui lui fourniraient des données d'actualités considérées comme proches même si les lignes éditoriales sont différentes.

Avec l'exercice du droit exclusif en ordre dispersé, seuls les titulaires de droits maîtrisant le contrôle technique pour le faire et/ou disposant de contenus à forte valeur ajoutée peuvent espérer une rémunération. Les autres auraient donc de plus en plus de difficultés à accéder à ce marché. De plus, les remontées aux auteurs, personnes physiques devraient faire l'objet d'attentions particulières.

De l'autre côté, ces négociations pourraient ne profiter qu'aux grands **acteurs de l'IA** qui ont les moyens financiers, humains et administratifs de négocier, fermant là encore le marché aux fournisseurs et déployeurs d'IA plus modestes. Les opérateurs les plus importants disposent en effet d'avantages concurrentiels du fait de leur intégration verticale (*Avis 24-A-05 du 28 juin 2024*, 2024).

- En amont, pour la production de modèles, ils bénéficient d'un accès direct à une puissance de calcul et à une grande quantité de données associées à l'utilisation de leurs multiples services.

- En aval, ils disposent de canaux de diffusion préexistants pour distribuer leurs modèles ; d'une part des modèles sont vendus de manière complémentaire à la vente d'accès à leur infrastructure de calcul ou leurs services de cloud ; d'autre part, les modèles de fondation sont parfois intégrés à des produits ou services existants (moteurs de recherche, réseaux sociaux, suites bureautiques, smartphones) disposant déjà d'une large base d'utilisateurs ; des entreprises indépendantes de distribution des logiciels d'IA peuvent ainsi être pénalisées.

Des positions dominantes ou des formes de concurrence déloyale pourraient donc apparaître en raison de l'intégration verticale évoquée ou des coûts fixes élevés qui poussent à des situations de monopole naturel. Le marché de l'IA reste cependant, à ce stade, très compétitif comme le montre l'entrée de nouveaux acteurs de dimension plus modeste tel DeepSeek (cf partie 3). La situation, susceptible d'évoluer très vite, est donc suivie avec attention par les autorités de la concurrence en Europe.

2.2 Les transferts de valeur obligatoires

Pour éviter aux ayants droit de se lancer sur ce marché en ordre dispersé, des réponses de transferts obligatoires sont envisagées soit dans le cadre de la propriété intellectuelle, soit en dehors de ce cadre.

Dans le cadre du droit d'auteur, l'opportunité, la faisabilité juridique, notamment dans un cadre international, de mesures alternatives et les évolutions législatives éventuelles nécessaires feront l'objet d'approfondissements dans la partie dédiée du rapport. Nous nous contentons ici d'appréhender la question sous le seul volet économique. A l'heure actuelle, dans des situations

déterminées, le droit exclusif « complet » qui comprend le monopole d'autorisation et la rémunération (cf. partie 4) est parfois atténué. Dans certains cas, la faculté d'autoriser ou d'interdire se déplace, de solutions individuelles vers des solutions collectives (gestion collective, licence collective étendue…) mais subsiste. Ces cas font donc partie du cadre économique de **solutions marchandes de contractualisation des droits.**

Dans d'autres cas, au contraire la faculté d'autoriser ou d'interdire disparaît. Dans des cas marginaux, les deux prérogatives – autorisation et rémunération - disparaissent au profit d'une exception non compensée (exception pour parodie par exemple). Dans d'autres cas, seule la faculté d'autoriser ou d'interdire disparaît au profit d'un simple droit à rémunération (licence légale donnant droit à une rémunération équitable ou exception compensée pour la copie privée). Ce sont ces cas ou **la faculté d'autoriser ou d'interdire disparaît et la capacité de contractualisation sur une marché également au profit de solutions organisées sous la tutelle des Pouvoirs publics, que nous désignons sous le vocable « transferts obligatoires »**.

Les montants générés par la licence légale ou la copie privée sont loin d'être négligeables. Sur la base de ces exemples connus de droit à rémunération, des mécanismes de transferts de valeur obligatoires dont le cadre résulterait de discussions arbitrées par les Pouvoirs publics sont donc avancés en matière d'IA. France Digitale (*IA générative et droit d'auteur*, 2024) propose ainsi une exception compensée pour les contenus librement accessibles sur le web ; le droit exclusif continuant à s'appliquer pour les contenus « fermés ».

Les modalités de transfert obligatoire dans le cadre du DA présentent cependant des limites. Des éditeurs et des producteurs de presse et de musique soulignent que si le recours à ce type de transferts obligatoires permet certaines facilités, il comporte néanmoins des inconvénients tenant à un traitement uniforme des données protégées conduisant, en règle générale,à une contraction de la valeur des contenus les plus recherchés et de facto des rémunérations obtenues. Des producteurs audiovisuels notent qu'une rémunération compensatoire, qui n'intervient qu'a posteriori de la diffusion initiale de l'œuvre n'est pas une incitation à produire car elle ne permet pas de financer les œuvres, en particulier dans les économies nécessitant d'importants moyens financiers comme l'audiovisuel.

Ensuite, les réponses qui s'organiseraient sur une base purement territoriale ne répondraient pas à la mondialisation du marché de l'IA. De plus, l'articulation entre mesures obligatoires et libre jeu du marché reste un point essentiel à débattre (système alternatif ou cumulatif selon les secteurs ?). Enfin, les autorités publiques ne doivent pas contribuer à organiser une forme de concurrence déloyale au détriment des entreprises européennes vis-à-vis de leurs concurrents américains ou chinois, ni faire obstacle au développement des marchés en construction (cf. partie 3).

Ainsi, la mise en place d'une licence globale obligatoire, débattue puis rejetée il y a tout juste 20 ans en France, si elle pouvait apparaître à certains comme une solution juridique d'urgence rassurante, face à l'ampleur de la piraterie numérique, aurait sans doute mis en en grande difficulté la construction par la suite, de marchés d'offres légales. A l'instar de la licence globale, dans le cas de l'IA, les mécanismes de transferts obligatoires présentent les mêmes limites, la plus importante étant celle des incitations économiques ; la rémunération globale ne serait plus liée à l'attractivité et à la valorisation potentiellement croissante des œuvres avec l'IA mais à de longues négociations, à l'issue incertaine, et peu susceptibles d'évolutions rapides.

Cependant, la mise en place d'un marché des licences peinant à se généraliser à ce stade (voir supra), le débat ne peut être définitivement clos pour l'avenir. De tels mécanismes, avant toute mise en place, devraient être évalués sur le plan économique selon les défaillances de marché qu'ils résolvent par rapport à d'autres dispositifs (Lutes, 2025).

En dehors des évolutions dans le cadre de la PI, d'autres mécanismes sont envisageables. Mécanismes fiscaux tout d'abord comme des taxes affectées sur les chiffres d'affaires s'inspirant de modèles existants (taxes alimentant le compte de soutien du CNC…), obligations de financement selon la logique économique qui oblige l'aval (les diffuseurs) à financer l'amont (la création et la production) ou encore instauration d'un domaine public payant au profit de la création humaine. Ces pistes restent à explorer.

2.3 L'hypothèse d'une voie complémentaire : accompagner la construction d'une place de marché

Entre les limites de solutions marchandes en ordre dispersé et le manque de réactivité aux évolutions des marchés des transferts obligatoires, une voie complémentaire et optionnelle, celle d'un accompagnement collectif des conditions structurelles facilitant la construction d'une place de marché - c'est-à-dire d'un espace d'échange structuré permettant la contractualisation dans le respect des spécificités sectorielles - mérite d'être explorée.

Le rapport interministériel sur l'IA de 2024 proposait la mise en place d'une infrastructure technique pour les détenteurs de données patrimoniales appartenant au domaine public (Aghion & Bouverot, 2024). Les données issues d'activités culturelles, mobilisées par les systèmes d'intelligence artificielle rassemblent des objets aux statuts juridiques pluriels. Certaines de ces ressources sont détenues par des institutions culturelles publiques patrimoniales (bibliothèques, musées, archives …) et peuvent être, ou pas protégées au titre de la PLA. Depuis plusieurs années, le cadre réglementaire qui définit les modalités de mise à disposition et de réutilisation des informations publiques pour les personnes publiques (l'État, les collectivités territoriales, les personnes morales de droit public) et les personnes morales de droit privé chargées d'une mission de service public, a en effet beaucoup évolué. L'ouverture aux usages numériques de ce vaste patrimoine, qui peut correspondre à des objectifs de politique publique de rayonnement de la culture française, pose des questions spécifiques qui ne sont pas celles traitées par cette mission ; celle-ci se concentre sur les données protégées par la PI et détenues par des opérateurs privés.

Pour ces dernières données, l'objectif d'une place de marché commune est de **regrouper,** pour des catalogues ou des pans de catalogues définis, **une triple activité d'accès/d'autorisation/de rémunération**. La place de marché rassemblerait en effet, dans un même espace numérique, non seulement une infrastructure technique de mise à disposition des fichiers, mais également les autorisations juridiques d'exploitation et les conditions économiques de rémunération.

Afin de respecter les spécificités des différents secteurs culturels et des différents acteurs, la participation se ferait uniquement sur la base du volontariat. Il ne s'agit pas d'instaurer une nouvelle obligation légale et en aucun cas cette participation ou absence de participation, ne pourrait se substituer au respect des législations nationales ou régionales existantes ni à l'organisation de la gestion des droits, différente selon les secteurs. Certains secteurs sont en effet familiers de systèmes de gestion collective (volontaires ou obligatoires) ; d'autres préfèrent des négociations individuelles pour exercer leurs DPI. Certains secteurs comme ceux de l'image ne disposent pas de bases de métadonnées complètes agrégées. La finalité n'est pas de créer ex nihilo un dispositif totalement nouveau mais de s'appuyer sur les expertises et compétences reconnues et les missions variées des différents acteurs en place (organismes de gestion collective, éditeurs, producteurs) pour proposer aux opérateurs d'IA une offre intégrée.

Des grandes institutions comme la Bibliothèque nationale de France (BNF) ou l'Institut national de l'audiovisuel (INA) pourraient jouer un rôle dans le dispositif. Ces institutions mettent actuellement à la disposition de tiers, des données-œuvres encore protégées, avec l'accord des auteurs ou ayants droit (qui peuvent alors le cas échéant négocier les conditions juridiques et financières d'exploitation avec le tiers-demandeur). Ce rôle pourrait être étendu à des situations

évoquées ici. La BNF conserve, par exemple, un corpus de données-œuvres conséquent, notamment dans le domaine de l'écrit, susceptible d'intéresser de nombreux acteurs développant des systèmes d'intelligence artificielle. Sans pouvoir délivrer les autorisations ou rémunérer les auteurs, ces institutions, comme d'autres acteurs, pourraient jouer un rôle d'agrégateur afin de livrer, à l'échelle, des fonds numérisés sur lesquels elles disposent d'une expertise technique reconnue.

Sur la place de marché, des catalogues seraient mis à disposition dans cet espace commun, soit par des organisations disposant déjà de catalogues constitués (OGC, éditeurs, producteurs …), soit par de nouveaux intermédiaires techniques, tandis que les fournisseurs d'IA bénéficieraient d'un accès facilité à des données – œuvres de qualité. Dans tous les cas, les acteurs concernés conservent leur faculté de négocier et de fixer les prix. La détermination des prix serait notamment différenciée selon la destination des activités d'IA, sur la base de licences négociées annuellement (cf. partie 4).

En aucun cas, l'existence d'une place de marché ne peut empêcher des acteurs de contractualiser en dehors de cette place s'ils le souhaitent ou de ne pas contractualiser s'ils préfèrent s'opposer à des utilisations par l'IA, en mobilisant pour cela les prérogatives juridiques à leur disposition. L'objectif est de créer des incitations pour encourager les acteurs intéressés à s'associer à une opportunité économique.

Les avantages attendus sont nombreux.

Assurer les **conditions économiques** de retour sur investissements dans la création humaine. Comme nous l'avons noté (cf. partie 1), l'investissement dans des données-œuvres humaines de qualité, reflétant la diversité du monde réel, la diversité des langues, des cultures et des régions du globe apparait nécessaire à la fois sur le plan culturel et sur celui de l'innovation dans l'IA.

Eviter les risques mutuels pour les Parties, d'insécurité juridique associée à une potentielle utilisation frauduleuse d'œuvres protégées en centralisant les informations sur les droits et les personnes habilitées à autoriser. Dans un marché mondialisé de l'IA, où les entreprises sont établies dans des territoires aux législations diverses en matière de propriété intellectuelle, en confiant aux Parties la négociation des droits accordés, les contrats signés peuvent d'emblée avoir une portée mondiale ou régionale, contrairement aux réglementations lorsqu'elles restent purement nationales. Notons que la portée internationale de l'IA act pour l'entraînement des modèles d'IAG commercialisés dans l'UE et les conflits de lois éventuels, feront l'objet de travaux ultérieurs spécifiques du CSPLA. De plus, **l'insécurité juridique est porteuse de risques économiques majeurs** pour des sociétés d'IA qui pourraient être soumises à des décisions judiciaires douloureuses sur le plan financier, notamment des entreprises européennes qui ont un besoin crucial de lever des fonds pour se développer.

Techniquement, limiter les coûts de transaction liés à la recherche de données grâce à un accès simplifié, mutualisé et le plus automatisé possible à des données culturelles organisées, bénéficiant d'une garantie de fiabilité, de qualité et de diversité. Certains des opérateurs de l'IA ont mis en avant la question de l'accessibilité technique aux fichiers ; les acteurs qui sont en capacité d'offrir des autorisations juridiques, étant souvent à la fois pluriels pour un même objet et différents de ceux qui disposent des fichiers de contenus et des métadonnées associées. D'où la place envisagée d'intermédiaires techniques lorsque les titulaires de droits ne sont pas en capacité d'exercer ce rôle.

Enfin, sur le plan de **la diversité,** permettre aux titulaires de droits les plus modestes ou disposant de catalogues pointus, ou à leurs représentants, en bénéficiant d'une infrastructure collective, de rendre accessibles ces catalogues aux entreprises de l'IA intéressées. En miroir, permettre aux acteurs de l'IA de dimension modeste qui ne disposent pas de la taille critique

pour mobiliser les services adéquats de développer des offres diversifiées et ouvrir ainsi le marché aux capacités d'innovation des nouveaux entrants. Enfin, en ouvrant cette place de marché à l'ensemble des acteurs, y compris extra-européens qui le souhaitent, cet espace favoriserait la promotion de cultures minoritaires, conformément à une tradition française qui s'enorgueillit de soutenir des œuvres et des créateurs du monde entier.

Rendre opérationnelle cette solution nécessite, à n'en pas douter, qu'un certain nombre de sujets sensibles soient débattus et décidés entre les parties prenantes en particulier concernant les règles de gouvernance et de financement ; l'infrastructure nécessaire et la mise à disposition de données de qualité entraîne en effet mécaniquement des coûts dont la prise en charge doit être discutée et répartie. A l'issue du sommet de l'IA a notamment été annoncée la création d'une nouvelle Fondation issue de fonds publics et privés ; l'articulation de la place de marché avec d'autres initiatives reste donc à imaginer.

***

Dans les conditions actuelles, en France comme dans de nombreux autres pays, les négociations directes entre les acteurs peinent à se mettre spontanément en place. C'est pourquoi la mission propose un dispositif de place de marché qui joue le rôle d'accélérateur de ces négociations. De telles solutions de marché n'excluent pas l'opportunité pour des situations spécifiques, d'autres dispositifs juridiques plus contraignants dans le cadre ou hors du cadre du droit d'auteur.

3 – Chaîne de valeur et acteurs des systèmes d'IA

3.1 Typologie des systèmes et modèles

Pour comprendre la chaîne de valeur, on distingue usuellement **modèles d'IA et système d'IA.** Les modèles d'IA sont intégrés dans des systèmes d'IA dont ils sont une composante essentielle. Les modèles d'IA nécessitent l'ajout d'autres composants tels **qu'une interface utilisateur en aval** ou **des services de cloud en amont** pour devenir des systèmes d'IA. Un système d'IA est donc un système automatisé qui est conçu pour fonctionner à différents niveaux d'autonomie et peut faire preuve d'une capacité d'adaptation après son déploiement. Il déduit, à partir des entrées qu'il reçoit, la manière de générer des sorties telles que des prédictions, du contenu, des recommandations ou des décisions qui peuvent influencer les environnements physiques ou virtuels. Un système d'IA à usage général a la capacité de répondre à diverses finalités, tant pour une utilisation directe que pour une intégration dans d'autres systèmes d'IA ; il est fondé sur un modèle d'IA à usage général.

Au sein des modèles, on distingue en effet les modèles de fondation, à usage général, et les modèles spécialisés. Un modèle de fondation est un modèle de grande taille, pré-entraîné sur d'énormes quantités de données non étiquetées. Il est conçu pour être adapté à un large éventail de tâches différentes notamment après un ajustement (*fine-tuning*) supplémentaire (voir infra). Par ailleurs, des modèles d'emblée spécialisés sont développés pour des tâches spécifiques, sans passer par l'ajustement des modèles de fondation. Un modèle spécialisé pour un domaine n'est pas toujours de petite taille ; ainsi BloombergGPT, un modèle financier de Bloomberg comprend 50 milliards de paramètres.

Enfin, les modèles d'IA **générative** (*AI generative* ou *AI gen*) ont la capacité de générer du texte, des images et des vidéos à partir d'instructions textuelles. Les applications courantes impliquent des utilisateurs saisissant des instructions en langage naturel pour générer des résultats. Tous les modèles génératifs ne sont pas nécessairement des modèles de fondation. Les modèles de fondation sont plus larges dans leur conception et leur application potentielle. Ils peuvent être adaptés à des tâches non génératives comme la classification ou l'analyse.

Une partie de l'IA générative repose sur des grands modèles de langage (*large language model*, LLM). Le modèle génère la réponse la plus probable à une suite de mots produite par l'utilisateur (un prompt, une requête). Une autre partie repose sur **les modèles de diffusion**, typiquement utilisés dans la génération d'**images** par un utilisateur via un prompt. Ces réseaux de neurones sont dits « grands » compte tenu du nombre de leur paramètres : GPT-3, par exemple (utilisé par Open AI jusqu'il y a peu) comprend 175 milliards de paramètres. La quantité de données dans un texte étant énorme, il faut que le LLM comprenne un très grand nombre de paramètres. Concrètement, les LLM sont entraînés sur de grands ensembles de données textuelles comme Common Crawl8, The Pile9, MassiveText10, Wikipedia ou GitHub. Ces ensembles de données contiennent jusqu'à 10 000 milliards de mots ce qui est très couteux en ressources informatiques et en temps. Les modèles **multi- modaux, quant à eux** associent en toute fluidité texte, images, et paroles. Des entreprises comme Runway (qui compte parmi ces investisseurs Google et Nvidia)(Elias, 2024) ou Synthesia monétisent ainsi des solutions de création automatisée de vidéos en s'entraînant sur des textes, des images et des vidéos, auprès de clients qui utilisent par exemple ces vidéos à des fins marketing.

3.2 Les segments de la chaine de valeur

Dans l'AI act (article 83), le « **fournisseur** » est une personne physique ou morale, une autorité publique, une agence ou tout autre organisme qui **développe** ou fait développer **un système d'IA** ou un modèle d'IA à usage général et le met sur le marché ou met le système d'IA en service sous son propre nom ou sa propre marque, à titre onéreux ou gratuit. L'activité comprend la gestion des ressources en amont et le développement (European Commission, 2021).

Le « **déployeur** » est une personne physique ou morale, une autorité publique, une agence ou un autre organisme **utilisant** sous sa propre autorité un **système d'IA** sauf lorsque ce système est utilisé dans le cadre d'une activité personnelle à caractère non professionnel. Ce sont les utilisateurs. On peut citer parmi eux des hôpitaux et établissements de santé utilisant l'IA pour le diagnostic, des institutions financières utilisant l'IA pour l'évaluation des risques, des entreprises de recrutement utilisant l'IA pour le tri des CV, etc.

La chaîne de valeur d'un système d'IA peut être plus précisément segmentée en trois principaux blocs (Hoppner & Streatfeild, 2023).

- **Les ressources.** *Calcul.* En amont, les développeurs de l'IA ont besoin, outre d'une main d'œuvre qualifiée et de données en quantité, d'offrir la puissance de calcul et de stockage nécessaires. Pour y répondre, une entreprise peut acheter directement des capacités de calculs, les louer (services de cloud) ou utiliser des infrastructures techniques existantes. Les calculs se font principalement sur des puces, en l'occurrence des processeurs graphiques GPU (*graphics processing unit*) dont le marché est concentré autour de l'entreprise US Nvidia qui détient en 2023 85% des parts de marche des GPU dans le monde ou les GPUs développés par Google. Les services de cloud, permettant le stockage sur des serveurs (plutôt que sur un seul ordinateur), des modèles et des données, sont dominés par des services américains AWS (Amazon), Azure (Microsoft), GCP (Google). *Données.* Des services proposent la création, la collecte ou la préparation des données pour entraîner les modèles d'IA.

- La **modélisation** correspond au **développement** et à l'**entraînement** des modèles, en particulier des larges modèles de fondation entraînés sur de gigantesques quantités de données. Ces modèles sont fermés ou accessibles en open source. Certaines entreprises sont spécialisées dans le stockage et le partage de modèles ouverts.

- En aval, les modèles sont **déployés et commercialisés** en direction des services et utilisateurs finaux, via des applications développées par les développeurs de modèles ou par une tierce partie.

3.2.1 Les ressources –données

L'accès aux données s'effectue à partir de différentes sources (Figure 4) :

- Données disponibles publiquement. Les données issues du *web scraping* (protégées ou pas par la PI, obtenues ou pas de manière licite) et les ensembles de données open-source continuent d'être importantes pour le développement des modèles. Ce type de données constitue la majeure partie des données utilisées pour le pré-entraînement des modèles de fondation.

- Données synthétiques. Plusieurs modèles de fondation récemment publiés ont utilisé des données synthétiques. Outre la réduction des coûts, ces données répondent à des préoccupations de confidentialité/protection des données. Il existe cependant des

limites à leur utilisation en raison du risque d'effondrement des modèles à long terme (cf. Partie 1).

- Données propriétaires de tiers. Ces données sont collectées par des entités externes comme des courtiers de données.

- Données directement propriétaires. Ces données, détenues par des entreprises actives dans le développement de modèles de fondation, peuvent ne pas être accessibles à leurs concurrents (cf. Partie 2).

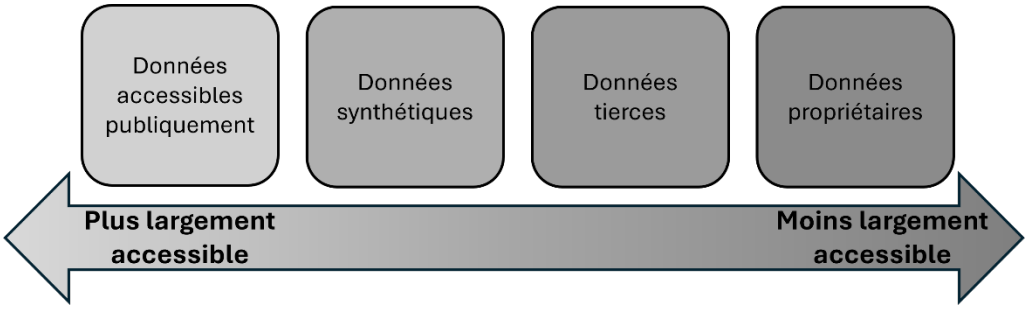


Figure 4. Les différent jeux de données et leur accessibilité (AI Foundation Models, 2024)

3.2.2 Le développement : les étapes de la modélisation

Le développement d'un modèle d'IA repose schématiquement sur deux grandes étapes successives. L'usage des données varie à chaque étape.

a) Une phase d'entraînement

Le **pré-entraînement** tout d'abord, se réfère à l'entrainement d'un modèle de fondation au cours duquel quatre tâches différentes sont réalisées :

1) La collecte des données d'apprentissage appelées données d'entraînement (en général, plus le modèle est complexe, plus il est performant et plus il nécessite un large jeu de données) ;
2) La construction d'un modèle (sous la forme d'un réseau de neurones, qui comprend une couche d'entrée, des couches intermédiaires et une couche de sortie) pouvant potentiellement relier les données d'**entrées** ou *input* (e.g. des images d'animaux) et de sorties désirées (les différentes espèces d'animaux) ; au départ, les paramètres prennent des valeurs initiales aléatoires et le modèle fournit une réponse (aléatoire) appelée **sortie** ou *output* (e.g. une image de chat est labelisée comme étant celle d'un chien) ;
3) La définition d'une fonction de coût à minimiser ;
4) L'entraînement, c'est-à-dire détermination de la valeur des paramètres qui minimisent la fonction de coût de façon à faire converger les sorties du modèle et les sorties désirées (e.g. une image de chat est correctement labellisée comme « chat », etc.).

D'autres étapes peuvent être implémentées. Le **réglage fin, ou ajustement, ou affinage** (*fine-*tuning) consiste **à spécialiser le modèle de fondation, en le « réentraînant » sur des données ou des tâches spécifiques.** Par exemple, le traitement de textes spécifiques (un rapport financier, un texte de loi) ou une tâche spécifique (analyse de sentiments, reconnaissance de termes

« métiers », détection de pièce défectueuses). Cela peut être fait en modifiant une partie des paramètres du modèle, mais seulement à la marge (pour ne pas effacer l'entraînement), et potentiellement en ajoutant une ou plusieurs couches au modèle (pré-)entraîné.

Le plus souvent, le modèle est entraîné, sur une partie seulement des données, les **données d'entraînement** (par exemple 80% des données) qu'on va chercher à expliquer le mieux possible en minimisant la fonction de cout, sous la contrainte de minimiser l'erreur de généralisation quand on essaie de prédire le reste des données (par exemple, 10%), appelées **données de validation**. Toutefois, les données de validation ne sont pas un test parfait de la capacité du modèle à généraliser, puisqu'elles participent elles-mêmes à l'entraînement du modèle. Le modèle pourrait alors échouer à prédire correctement de nouvelles données. *In fine*, on utilise souvent les 10 % restant pour tester le modèle sur sa capacité à généraliser à des données encore jamais vues, ce que l'on appelle **les données de test,** et donner une indication de performance du modèle.

b) Une phase d'inférence, de mise en production

Après la phase d'entraînement, l'inférence correspond à l'opération de mise en production du modèle c'est-à-dire du processus par lequel un modèle entraîné préalablement, va produire un résultat, des prédictions sur de nouvelles données. L'inférence peut être complétée par l'apport de nouvelles données, de données fraiches, pour que le modèle fournisse des informations en tenant compte de l'actualité ou de données bien spécifiques que le modèle va chercher dans une source externe. Le modèle n'est pas entraîné sur ces données. C'est ce que l'on nomme la génération augmentée de récupération (*Retrieval Augmented Generation* (RAG)ou encore parfois l'ancrage du modèle (*grounding*), qui ne relève pas de l'entrainement.

L'ancrage fournit des informations externes au modèle lors de son utilisation et n'altère pas ses paramètres internes. L'outil externe, auquel on peut faire appel, le plus évident est un navigateur web, permettant au modèle de rester à jour**,** mais les praticiens affinent les modèles de langage pour qu'ils puissent utiliser des appels d'API (*Application Programming Interface*, correspondant à une interface entre logiciels, voir infra) et accéder ainsi à une grande variété d'outils. Lorsqu'une requête est reçue, le système effectue une recherche dans un ensemble de documents ou de données externes. Les informations pertinentes sont récupérées et utilisées pour enrichir le contexte de la génération. Le modèle de langage génère ensuite une réponse en s'appuyant à la fois sur ces informations externes et sur ses connaissances internes. Cela améliore la précision et la pertinence des réponses générées, permet d'utiliser des informations à jour sans nécessiter un réentraînement du modèle, et réduit le risque d'"hallucinations" en ancrant les réponses dans des faits vérifiables. L'ancrage est particulièrement utile pour les entreprises souhaitant utiliser des modèles de langage sur des faits comme l'actualité.

Pour résumer, trois grandes catégories de données sont utilisées dans le développement des modèles :

1) Des données d'entraînement, pléthoriques, de l'ordre de plusieurs millions ou plusieurs milliards ;
2) Des données d'affinage, spécialisées, qui peuvent être disponibles sur internet ou sélectionnées avec soin par une entreprise ou une organisation ;
3) Des données « fraîches » consistant à ancrer le modèle dans l'actualité sans nécessiter d'entrainement.

L'influence d'un jeu de donnée n'est pas la même à chaque étape. Enlever un jeu de données dans la phase d'entraînement n'influence que faiblement la performance du modèle tant il est entraîné sur une quantité immense de données. Pour l'affinage, un jeu de données pertinent pour l'usage du modèle est crucial, tandis que si le jeu de donnée ne l'est pas, sa valeur est nulle. Il en va de même pour l'ancrage. De plus, ce ne sont pas toujours les mêmes acteurs qui procèdent aux différentes tâches en s'appuyant soit sur la quantité des données soit sur leur qualité. Enfin, pour

les start-ups d'IA dont les produits sont orientés vers une demande bien spécifique, le coût principal n'est pas l'entraînement mais l'accès à des modèles déjà entraînés.

3.2.3 Lancement des modèles et déploiement

Une fois les modèles entrainés, ils sont publiés, c'est-à-dire qu'ils sont disponibles au déploiement.

Le modèle est alors disponible en open source ou dans un format propriétaire. En open source, il est potentiellement utilisable sur de nouvelles infrastructures et peut être étudié et modifié. C'est le cas –dans une certaine mesure– de LLaMA (Meta) ou des modèles du Français Mistral AI. Dans les modèles propriétaires au contraire, comme GPT (Open AI), Gemini (Google), Claude 3 (Anthropic), l'accès est contrôlé par des licences, des *plugs-in*, des API, etc.

Une fois publiés, les modèles d'IA peuvent être affinés, non plus par le développeur initial (voir supra) mais par un tiers, utilisateur, entreprise ou intermédiaire (comme Eviden) qui agit sur demande d'un client. Ces affinages peuvent être intégrés à des logiciels et des applications. Outre les outils de fine-tuning disponibles sur les plateformes de développement de modèles de fondation (FM, *Fondation Models*), tels que ceux proposés par Microsoft, Amazon et Google, certaines entreprises, dont OpenAI et Mosaic, offrent des services de *fine-tuning*. Ces services peuvent être utiles pour les déployeurs et les clients qui ne disposent pas des ressources techniques internes pour développer des modèles de fondation, mais qui souhaiteraient tirer parti des vastes capacités des FM tout en bénéficiant d'une solution personnalisée (qui peut inclure un mélange de modèles ou de points d'accès, comme des API).

Il existe une variété d'options pour accéder aux modèles de fondation et les déployer. Les entreprises peuvent choisir la manière dont elles accèdent aux modèles, depuis l'accès API des modèles *open-source* jusqu'au développement de leur propre modèle. En fonction de leurs besoins et de leurs contraintes financières, des entreprises peuvent inclure un mélange de modèles les plus puissants (utilisés via des API, moyennant finance ou non) et de modèles moins performants (par exemple maisons), ou un mélange de modèles propriétaires et open-source.

Les modèles sont mis à la disposition des utilisateurs via une application (sous forme de chat, comme ChatGPT de la société OpenAI ou Le Chat de Mistral AI) ou via une interface de programmation pour les développeurs (API), qui permet à un programme informatique une interaction directe avec le modèle. L'accès via une API consiste à autoriser les utilisateurs à interagir (i.e. en envoyant des requêtes) avec le modèle stocké sur un serveur par le fournisseur. C'est le cas du service GPT d'OpenAI. Ces accès sont souvent payants avec des restrictions d'utilisation, des limites de requêtes ou des tarifs différenciés en fonction du volume de requêtes ou des fonctionnalités accédées. L'accès via des plateformes Cloud, comme Amazon Wed Services (AWS), Google Cloud ou Microsoft Azure correspond à l'hébergement de modèles d'IA accessibles aux utilisateurs.

Par ailleurs, il y a une augmentation de la disponibilité des accès API payants directement auprès des développeurs et via des plateformes de développement, telles qu'Amazon Bedrock. Par exemple, via Amazon Bedrock, les clients peuvent accéder à des modèles de fondation, notamment Claude 3 d'Anthropic, Command de Cohere, Jurassic-2 d'AI21, Llama 2 de Meta, Stable Diffusion de Stability AI, 8x7B de Mistral et Titan d'Amazon. Les clients peuvent également choisir de payer un accès API à certains de ces modèles directement auprès des développeurs, comme Anthropic, Mistral et Stability AI (*AI Foundation Models*, 2024).

Pour résumer, une entreprise peut choisir d'accéder aux modèles directement via le développeur ou au moyen d'une place de marché qui agrège différents modèles ; elle peut choisir d'utiliser un large modèle général, le plus souvent un modèle de fondation ou un plus petit modèle, souvent

spécialisé et développé maison ou via une entreprise spécialisée. Ces modèles peuvent être open-source ou propriétaires.

3.2.4 Usagers

Les résultats récents d'une enquête concernant l'utilisation de l'IA par les entreprises britanniques, issus de « Business insights and impact on the UK economy » de l'ONS, montrent que 15 % des entreprises britanniques utilisent actuellement au moins une des technologies d'IA mentionnées dans l'enquête (qui incluaient les services FM ainsi que d'autres technologies d'IA) ; ce chiffre atteint 46 % parmi les plus grandes entreprises (celles comptant 250 employés ou plus). L'adoption de l'IA a tendance à augmenter avec la taille de l'entreprise, et le niveau d'adoption varie considérablement selon les secteurs (*AI Foundation Models*, 2024; Business Insights and Conditions Survey Team, 2024). Les principales raisons pour lesquelles les entreprises utilisent l'IA sont : améliorer leurs opérations commerciales, fournir ou personnaliser un produit/service, développer un nouveau produit/service ou explorer un nouveau marché.

En dehors des entreprises, de nombreux usagers particuliers utilisent les modèles d'IA. Un sondage de 2023 mesure que 20% des adultes utilisent les IA pour créer un nouveau texte ou une nouvelle image et 36% utilisent ces modèles pour accéder à des informations (Online Nation 2023 Report, 2023).

3.2.5 Des circuits complexes de valorisation dans la phase de déploiement

La monétisation, par les fournisseurs de modèles de fondation des services auxquels ces modèles donnent accès est désormais quasiment systématique. Pour les clients entreprises, Amazon, Anthropic, Google, Microsoft, Mistral AI, OpenAI et Stability AI proposent une variété de niveaux payants (abonnements mensuels, paiements en fonction du nombre de *tokens* d'entrée et de sortie utilisés, paiement à l'utilisation, systèmes basés sur des crédits…). Pour les clients particuliers, la majorité des services, tels que ceux proposés par Google, OpenAI, Anthropic, xAI et Microsoft, sont soit gratuits, soit à hauteur de 20 $ par mois environ pour un abonnement.

Pour donner des ordres de grandeurs, en 2024, les revenus d'OpenAI proviennent d'une part des revenus API (estimés à 27%, 1 Md$), d'autres part des abonnements (73%, 2.7 Md$) (Khan, 2024; Sacra, 2024). Anthropic tire ses revenus essentiellement des API (60-75% de parties tierces (600-750 m$), 10-25% d'usage direct (100-250 m$) et environ 15% d'abonnements (150 m$) (Khan, 2024; Sacra, 2024).

On observe également une tendance à la monétisation des logiciels de productivité avec intégration de l'IA, présentée comme un complément premium ; le Copilot de Microsoft, par exemple, peut être acheté en tant que complément pour les utilisateurs existants de Windows Home, Pro ou Enterprise.

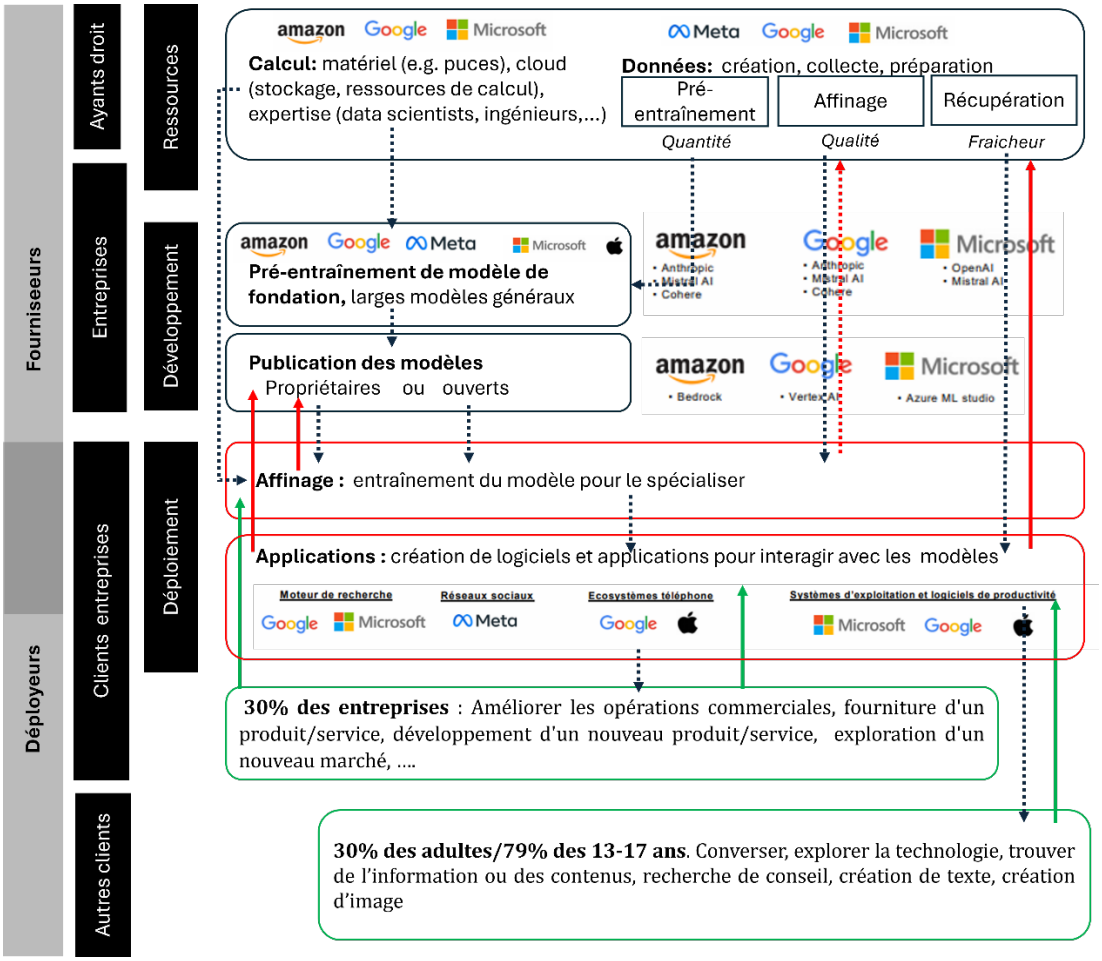


Figure 5. Aperçu de la chaîne de valeur des modèles de fondation (adapté de (AI Foundation Models, 2024; Trésor, 2024)). Les flèches noires indiquent l'usage de différentes composantes dans le développement des modèles de fondation. Les flèches rouges indiquent le paiement des opérateurs pour utiliser les modèles, l'affiner, ou l'alimenter. Les flèches vertes indiquent le paiement des utilisateurs.

Par ailleurs, **la monétisation passe également, pour d'autres opérateurs, par leurs activités d'affinage des modèles de fondation et la création d'interfaces utilisateur.** Moyennant paiement aux développeurs de modèles de fondation, des entreprises réalisent leurs propres applications qu'elles facturent à leur tour à leurs clients. L'affinement est parfois réalisé sur des jeux de données « métier », fournies par le client lui-même. Sous réserve que certains modèles de fondation restent accessibles en open-source, les applications à partir de ces modèles peuvent ne pas être très onéreuses. D'autres modèles économiques, s'appuient sur des abonnements coûteux accessibles essentiellement aux grandes entreprises (*AI Foundation Models*, 2024).

A titre d'exemple, dans la musique, la majorité des applications de génération de contenus, entrainées sur un large corpus de titres existants, sont destinées aux amateurs. Les entreprises Suno AI ou Udio AI proposent des fonctionnalités à des prix accessibles au grand public (de la gratuité pour des usages limités à quelques dizaines de $ par mois). Soundful, autre entreprise génératrice de musique, plus adaptée à des professionnels, pratique quant à elle des tarifs nettement plus élevés pour les entreprises.

3.3 Opérateurs et tendances du marché

3.3.1 L'essor des partenariats

L'autorité de la concurrence britannique a identifié plusieurs catégories d'organisation de la chaîne de valeur entre les entreprises (*AI Foundation Models*, 2024):

- Intégration verticale. La même entité est présente à différents niveaux de la chaîne de valeur des FM, que ce soit dans les intrants des FM (comme le calcul ou le matériel), le développement des FM, et leur déploiement.
- Partenariats dans la chaîne d'approvisionnement des FM. Des partenariats existent entre les développeurs de FM et les fournisseurs de services cloud (CSP).
- Chaînes de valeur dispersées. Une gamme d'entreprises opère à différents niveaux de la chaîne de valeur. Par exemple, une entreprise fournit les ressources de calcul, une autre développe les FM, et une entreprise distincte déploie ce modèle dans ses propres produits et services.

L'autorité met en évidence le développement continu des partenariats, des investissements et des accords stratégiques pour les modèles de fondation (*AI Foundation Models*, 2024). Les partenariats peuvent offrir des avantages significatifs pour les parties impliquées et conduire à une innovation accrue et à des gains d'efficacité (accéder à des ressources rares, amener leurs modèles sur le marché plus rapidement et à plus grande échelle…). Depuis 2019, plus de 90 partenariats entre les entreprises « GAMMAN » et les « partenaires » ont été relevés. Les entreprises GAMMA (Google, Amazon, Microsoft, Meta, Apple) et Nvidia (qui est le principal fournisseur de puces accélératrices d'IA) sont dénommées les « GAMMAN ». Les développeurs de modèles de fondation, les déployeurs de ces modèles ou les fournisseurs d'outils pour développeurs de ces modèles sont dénommés les « partenaires ». Par ailleurs, des start-ups de l'IA sont souvent rachetées par des acteurs plus importants.

L'autorité de la concurrence britannique observe également une grande variété de structures de partenariats (*AI Foundation Models*, 2024):

- **Partenariats de données** : Les partenariats peuvent permettre à une partie d'accéder aux données de l'autre partie (Meta et Shutterstock, Google et Reddit,…) ;

- **Partenariats de calcul** permettent aux développeurs de modèles de fondation d'accéder à des ressources de calcul, y compris l'accès à des systèmes spécialisés de supercalcul ou à des puces (Microsoft et OpenAI, Amazon et Anthropic, Google et Anthropic …) ;

- **Partenariats de distribution** : Ceux-ci peuvent prendre plusieurs formes.

    - **Distribution de modèles de fondation** : Certaines entreprises construisent des plateformes de développement offrant une bibliothèque de modèles de fondation. Les partenariats peuvent permettre à une entreprise GAMMAN (1) d'ajouter le ou les modèles du partenaire à sa bibliothèque ou (2) de fournir un accès au(x) modèle(s) du partenaire via les outils de développement de modèle de fondation de l'entreprise GAMMAN (Amazon et Cohere, Google et Mistral, Microsoft et Meta, Amazon et HuggingFace…).
    - **Distribution d'outils** : Les entreprises GAMMAN peuvent également ajouter l'outil de développement de modèle de fondation du partenaire à leur propre plateforme ou marketplace (Microsoft et Nvidia…).

- **Distribution de l'infrastructure de modèle de fondation.** Une entreprise GAMMAN peut distribuer l'infrastructure IA d'un partenaire via son propre marketplace cloud (Nvidia et Google, Nvidia et AWS…).
- **Programme d'accélérateur**. Les entreprises GAMMAN peuvent créer un programme d'accélérateur pour les start-ups partenaires en IA (le programme pour start-up Meta / Hugging Face / Scaleway). Ceux-ci peuvent offrir des financements, des ressources de calcul, ainsi que des opportunités de coaching et de réseautage.

- **Investissements** : Les entreprises GAMMAN peuvent être l'un des multiples investisseurs dans une entreprise partenaire, aux côtés d'autres entreprises GAMMAN. Les sociétés de capital-risque participent également couramment à ces tours de financement (levées de fonds pour Runway AI, Cohere, Adept, Inflection et HuggingFace, ….).

3.3.2 Un oligopole d'entreprises américaines domine le marché du développement des modèles de fondation

Selon ADLC, l'IA est la première technologie d'emblée dominée par des géants (*AI Foundation Models*, 2024). Certaines entreprises peuvent jouer plusieurs rôles dans la chaîne de valeur. Par exemple, Google peut être à la fois fournisseur de ses propres systèmes d'IA et utilisateur de technologies d'IA tierces dans ses produits.

Le marché du développement est dominé par des modèles de fondation américains. On compte parmi les fournisseurs Google (développeur de Gemini), OpenAI (créateur de ChatGPT et GPT), Meta (développeur de LLaMA), Microsoft (partenaire d'OpenAI), DeepMind (filiale de Google), Anthropic (développeur de Claude), Stability AI (créateur de Stable Diffusion), etc. (cf. Figure 6). Google, Microsoft et Meta représentent un quart des modèles de fondation récents sur les 348 modèles recensés (Trésor, 2024).

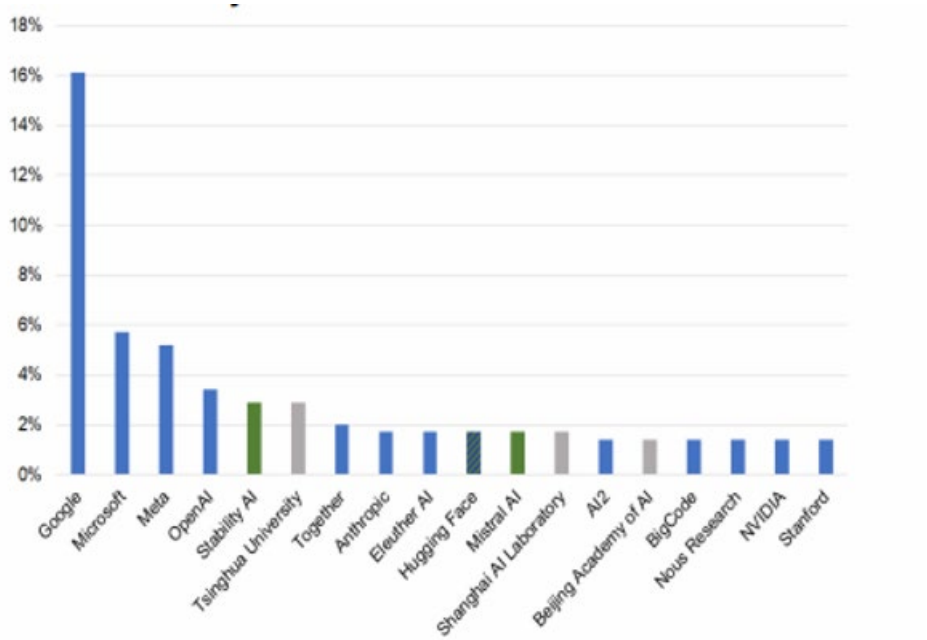


Figure 6. Part des différentes organisations dans le nombre total de modèles de fondation entre le 1er janvier 2022 et le 4 octobre 2024. (Trésor, 2024)

3.3.3 Small is beautiful

Bien que les modèles de fondation actuels continuent d'augmenter en taille, il est peu probable que de nombreux acteurs viennent rejoindre le club restreint des entreprises qui les fournissent, compte tenu du coût de développement en ressources à la fois économiques et environnementales. D'autant plus que face à des coûts élevés, les revenus potentiels sont parfois aléatoires et l'équilibre économique incertain. C'est pourquoi, **en matière de développement, des tendances à la sobriété, à la spécialisation et à l'hybridation** se dessinent.

L'entreprise Perplexity AI, propose ainsi un moteur de recherche basé sur des modèles d'IA « hybrides », c'est-à-dire sur des modèles de fondation proposés par d'autres entreprises mais aussi sur son propre modèle et sur la RAG. De plus, il existe un intérêt croissant pour le développement de modèles plus compacts (avec un nombre réduit de paramètres) qui, tout en offrant des capacités étendues, nécessitent moins de ressources pour leur développement ou leur déploiement. Cette tendance est en partie motivée par les coûts de calcul et par le fait que de nombreux cas d'utilisation ne requièrent pas la pleine capacité des grands modèles à usage général. Les nouveaux entrants Mistral et Deepseek, bien que très différents, ont fait de cette sobriété, un argument marketing. L'agent conversationnel Deepseek, issu d'une start-up chinoise début 2025, est un LLM généraliste, partiellement en open-source, se positionnant pour concurrencer OpenAI ; le système se démarque des autres parce qu'il fonctionne avec des GPU de capacités réduites(Laird, 2025; Nellis & read, 2025). Son coût de développement, tel qu'annoncé par les autorités chinoises, serait par ailleurs très inférieur à celui des géants américains. Les modèles du fleuron français Mistral sont quant à eux plus petits que ceux des géants (modèle « Mixtral 8x7B » comptant 46,7 milliards de paramètres mais n'en utilisant que 12,9 milliards par *token (*un *token* est une unité correspondant à un groupe de lettres). L'entreprise française publie aussi plusieurs petits modèles (entre 3 et 8 milliards de paramètres). Parmi les exemples récents de petits modèles, on trouve également Gemma 7B de Google et Zephyr 7B de Hugging Face, que les entreprises affirment être à la hauteur, voire surpasser, des modèles beaucoup plus grands sur certains critères. De plus, la « distillation » permet de réduire la taille d'un modèle tout en améliorant ses performances pour des tâches spécifiques ; elle consiste à créer un petit modèle efficient qui apprend à imiter un plus grand modèle en essayant de limiter la perte de performance. Le petit modèle est entraîné sur les prédictions du grand modèle plutôt que sur des données elle mêmes (*A Three-Step Design Pattern for Specializing LLMs*, n.d.; *AI Foundation Models*, 2024; *Distilling Step-by-Step*, n.d.)

Enfin, la spécialisation s'accroît, soit par le développement de modèles d'emblée spécialisés, soit par le ré-entraînement de modèles de fondation via le *fine tuning* ou encore par l'ancrage des modèles grâce à des données fraîches. Phi-2 de Microsoft, est un « petit modèle linguistique » comptant seulement 2,7 milliards de paramètres (Hughes, 2023). Développé avec des données d'entraînement soigneusement sélectionnées, Microsoft revendique des performances supérieures à celles du Llama 2 70B sur des critères spécifiques, tels que le codage ou le raisonnement de bon sens. Un autre modèle, dédié aux mathématiques, récemment publié est Orca-Math de Microsoft, créé en affinant le modèle Mistral 7B(*AI Foundation Models*, 2024; Hughes, 2024).

***

Ce bref panorama de la chaîne de valeur a pour ambition de mieux comprendre la *création* de la valeur sur différents marchés, afin de poser ensuite la question du *partage* de cette valeur au profit de la culture. Cette question de la création de valeur va bien au-delà de la phase de pré-entraînement des modèles de fondation des entreprises les plus médiatisées. L'extraction en masse de données directement accessibles sur le web est une méthode de collecte principalement

utilisée par les modèles de fondation, américains pour la plupart. D'autres données acquièrent leur valeur de part les usages spécialisés auxquels elles sont destinées.

Les différentes activités de pré- entraînement, de fine tuning ou d'ancrage sont parfois menées par les mêmes entreprises, parfois au contraire par des entreprises différentes. Dans un écosystème en construction, dont il reste à approfondir les contours, les *fournisseurs de modèles de fondation* monétisent les services d'application ou d'affinage auxquels ces modèles donnent accès auprès des usagers finaux (entreprises et particuliers). Par ailleurs, *d'autres opérateurs intermédiaires,* monétisent leurs propres activités d'affinage et de systèmes applicatifs les plus divers (cf. *Figure 5*).

La valeur des données culturelles dans cet écosystème global doit donc être envisagée à l'aune de ce constat. Certaines entreprises, pour exercer leurs activités commerciales, ont un besoin impératif de données culturelles de qualité pour différentes actions, qui ne se limitent pas à la fouille, afin d'apporter spécialisation et fraîcheur aux résultats proposés. L'adéquation des données disponibles avec les applications et usages multiples déployés par les opérateurs d'IA augmente la valeur de ces données et donc les rémunérations attendues. Ce ne sont donc pas seulement l'ensemble des acteurs mais aussi l'ensemble des types d'activités créatrices de valeur sur les marchés pertinents qui doivent fournir le socle du partage.

4 – Valorisation des données-œuvres pour les systèmes d'IA

Quel que soit le cadre dans lequel s'organisent les transferts (cf. partie 2) et compte tenu de la complexité de la chaîne de création de valeur dans les systèmes d'IA (cf. partie 3), la question de la valeur des données - œuvres se pose ainsi que celle du partage qui en découle.

4.1 Quantifier la valeur des données

Des recherches scientifiques récentes de quantification de la valeur des données pour les modèles d'IA permettent d'éclairer le débat. Une abondante bien que jeune littérature consacrée à la notion d'attribution des données consiste en effet à calculer la contribution marginale de chaque jeu de données à la performance du modèle en général et à la genèse d'un résultat particulier (une sortie) à la suite de la requête d'un utilisateur. Un point central dans ce cas est celui de la substituabilité de ce jeu de données, qui se décline en deux questions voisines : le modèle peut-il performer de la même manière, c'est-à-dire répondre correctement aux diverses requêtes qui lui sont formulées, sans ce jeu de données ? Quelle est la contribution relative d'un jeu de données à une sortie spécifique ?

Pour répondre à ces deux questions, dans la littérature, trois approches co-existent. La première consiste en des changements de paramètres des modèles (que ce soit pour entraîner les modèles sur des sous-ensembles de données, ou en altérant les paramètres du modèle déjà entraîné) afin d'établir des liens de causalité. La seconde, corrélationnelle, cherche à mesurer la similarité entre le résultat généré par le modèle et les éléments constituant la base de données d'entraînement. La troisième, causale et proactive (qui ne peut s'appliquer aux modèles déjà entraînés) correspond au marquage par filigrane des données ingérées.

4.1.1 Etablir des liens de causalité en modifiant les paramètres des modèles

La méthode de validation croisée par suppression d'une donnée (*leave-one-out*)

Une manière d'étudier l'influence d'un jeu de données est de l'entraîner sans ce jeu de données, et de comparer le contenu généré par l'IA entre le modèle entraîné sur le jeu de données entier et celui entraîné sur le jeu de données réduit. C'est la méthode de la validation croisée par suppression d'un point et dite du « *leave-one-out* ». Formellement, ce problème est souvent formulé comme suit : comment la suppression d'un point de données particulier de l'ensemble de données d'entraînement et le réentraînement subséquent du modèle affectent-ils sa sortie ? Ce changement dans la sortie sert de mesure de l'influence du point de données supprimé sur cette sortie spécifique du modèle. Cela se fait donc en deux étapes : l'entraînement sur le jeu réduit et une mesure quantifiant la différence de qualité entre les deux outputs (Maleki et al., 2014).

Les principales études visant à tester la faisabilité de cette approche génèrent tout d'abord des contenus avec les différentes manières d'amputer le jeu de données d'entraînement. Ainsi, l'influence réelle peut être connue, puisqu'on sait quel jeu de données a servi à générer un type de contenu. Ensuite, elle utilise la méthode d'amputation pour voir si on recouvre le bon jeu de donnée**,** c'est-à-dire, celui qui a effectivement servi à la génération de l'output. En procédant de la sorte, on peut corréler les résultats du test à la « réalité » et voir la qualité de la méthode. Ces études montrent des résultats significatifs mais le degré de corrélation entre les variables reste faible.

La méthode du *leave-one-out* comprend plusieurs limitations. Tout d'abord elle examine uniquement les conséquences de la suppression d'une source de données de l'ensemble du jeu

complet. Cette approche pourrait ne pas refléter avec précision l'importance d'un point de données en raison des interactions complexes potentielles entre les sources. Par ailleurs, les points de données dupliqués, courants dans de nombreux ensembles de données d'apprentissage automatique, peuvent ne pas créditer la contribution d'un item. Par exemple, considérons deux propriétaires de jeux de données aux caractéristiques presque identiques. La suppression de l'un ou l'autre de l'ensemble de données entraînerait probablement un changement minimal dans la probabilité de génération de contenu du modèle, rendant ainsi les scores *leave-one-out* de chacun proches de zéro. Ce scénario pourrait injustement ne pas attribuer de valeur à aucun des contributeurs, malgré le rôle crucial de leurs ensembles de données dans les performances du modèle. De plus, dans des situations avec de nombreuses sources de données, le score *leave-one-out* pourrait diminuer jusqu'à presque zéro, ne reconnaissant pas les contributions subtiles des sources individuelles.

Enfin, l'étape d'entrainement sur des jeux réduits de données a un coût computationnel prohibitif (en termes de temps, de ressources informatique et énergétique) si l'on veut connaître l'influence de chaque jeu de données, puisqu'il faut entraîner un modèle sur chaque sous-type de données. Certaines méthodes existent pour réduire le coût de cette opération (Hammoudeh & Lowd, 2024), qui n'en demeure pas moins élevé.

La méthode de la « Shapley value »

Une alternative au *leave-one-out* est présentée notamment dans une étude récente (J. T. Wang, Deng, et al., 2024): la mesure de l'impact incrémental de l'incorporation d'une nouvelle source de donnée dans l'entraînement du modèle, en considérant toutes les autres combinaisons possibles. **La démarche est inversée par rapport au *leave-one-out* ; au lieu d'amputer le modèle de certains jeux de données, ceux-ci sont rajoutés par itérations successives.** On mesure ainsi l'impact incrémental de l'incorporation d'une nouvelle source de donnée dans l'entraînement du modèle, en considérant toutes les autres combinaisons possibles. Autrement dit, l'idée est d'entraîner tout d'abord le modèle sur une base de données de faible taille. Puis, le modèle est affiné (*fine-tuned*) avec l'ajout séquentiel, une à une, de nouvelles données. Chaque affinage produit un nouveau modèle, qui comporte un certain nombre de jeux de données.

On peut procéder ainsi dans différents ordres. Par exemple, si on considère trois jeux de données, A B et C, le modèle peut être entrainé sur le jeu A, puis affiné sur le jeu de données B, puis affiné sur le jeu de données C. Enfin, on peut répéter la démarche en considérant plusieurs ordres (e.g. B, puis C et A, etc.). Cela est plus précis que de l'entraîner sur seulement A, seulement B, seulement C, A et B, B et C, etc (cf. exemple fictif dans l'Encadré 3 en Annexes). Cette approche permet de dériver une métrique de la contribution marginale d'un jeu de donnée à la création d'un contenu dans les différents ordres, et de répartir la rémunération en fonction cette contribution marginale (cf. Encadré 4 Figure 1). La probabilité de générer un output spécifique après chaque étape d'ajustement donne une estimation de l'importance de chaque base de données.

Cette approche permet de calculer la « valeur de Shapley », concept issu de la théorie des jeux coopératifs qui conduit à répartir équitablement les récompenses ou les coûts entre les participants en fonction de leurs contributions individuelles à un résultat collectif. La méthode de la valeur de Shapley tient compte de l'impact incrémental de l'incorporation d'une source de données aux côtés de toutes les combinaisons possibles d'autres sources.

Cette méthode annoncée comme une alternative plus précise que le *leave-one-out* présente cependant les mêmes difficultés en termes de coûts. Pour éviter le coût rédhibitoire du calcul de la valeur de Shapley sur un grand jeu de données, différentes propositions existent pour l'approximer. En voici quelques exemples, parmi une littérature récente abondante.

Des méthodes tentent de déterminer l'influence d'une donnée sur une sortie. Une méthode appelée *Influence function* pour approximer la valeur de Shapley (Koh & Liang, 2017) consiste à calculer l'influence d'un jeu de données sur les paramètres d'un modèle en surpondérant un point de données sans réentraîner le modèle, plutôt qu'en le supprimant puis en réentraînant le modèle pour évaluer l'influence du point de donnée supprimé (*leave-one-out*). Cette méthode, est soit très imprécise, soit très couteuse (si elle est appliquée à chaque sous-ensemble de modèles) (Jia et al., 2019) (Koh & Liang, 2017). Une autre méthode consiste à simplifier le modèle en diminuant son nombre de paramètres et en simplifiant le lien entre les entrées et leur sortie, afin de calculer l'influence de chaque point de données sur la sortie sans entraîner les modèles (ou seulement quelques-uns ; méthodes appelée TRAK). En utilisant cette approche, (J. Deng et al., 2024) montrent sur un jeu de données musicales de taille relativement modeste que cette méthode, rapide et donc peu couteuse à mettre en place à grande échelle, est corrélée avec la méthode d'entrainement des modèles sur les sous-ensembles des données (à la Shapley), mais modestement : 30%. Cette approximation est donc peu précise.

D'autres approches consistent en l'utilisation de méthodes d'échantillonnage aléatoire ou « intelligent ». Une équipe menée par un chercheur de l'Université de Californie à Berkeley a comparé le coût de calcul de différentes méthodes (Jia et al., 2019). La première consiste en un échantillonnage des modèles entraînés sur des sous-ensemble de données (des données réduites), dans le but d'approximer le calcul exact de la valeur de Shapley. Cette méthode est précise, plus rapide que d'entraîner le modèle sur tous les sous-ensembles de données, mais demeure inexploitable tant elle demande de temps d'exécution. Une autre approche est inspirée du test de groupe, qui consiste en la détermination du test optimal pour déterminer si un objet est défectueux. Par exemple, si l'on souhaite déterminer quelle ampoule parmi six est défectueuse, une possibilité est de les tester individuellement, avec le risque de devoir tester cinq fois, alors que l'on pourrait tester les ampoules par groupe (d'abord deux groupes de trois, pour identifier le groupe qui contient l'ampoule défectueuse). Cela permet d'échantillonner de manière efficace. On peut appliquer ce principe en groupant les données pour identifier non la qualité défectueuse mais l'utilité des données pour une sortie d'un modèle, ou sa valeur de Shapley (cf. Encadré 4 Figure 2). Cette approche est relativement précise mais reste trop couteuse pour être déployable.

D'autres auteurs ont proposé de désapprendre au modèle la sortie générée par le modèle, et d'évaluer les images qui sont moins bien représentées dans le nouveau modèle. Cette approche permet de quantifier la contribution des données d'entraînement sans réentraînement (S.-Y. Wang et al., 2024). Regénérer les données sans ces images conduit à une image très différente de l'image générée initialement. Cette méthode a été utilisée seulement avec des modèles de génération d'image (non au texte) mais surtout rien n'indique à ce stade qu'elle puisse être utilisée sur de larges jeux de données.

Citons enfin une dernière approche permettant d'approximer la valeur de Shapley en calculant la contribution d'un point de données pendant l'entraînement du modèle(J. T. Wang, Mittal, et al., 2024). Alors que le calcul de la valeur de Shapley nécessite de réentraîner le modèle plusieurs fois avec différents sous-ensembles de données pour calculer les contributions marginales, le Data Shapley en cours d'exécution (*In-Run Data Shapley*) résout ce problème en exploitant la nature itérative des algorithmes d'entraînement, qui se fait donc par étapes. En effet, à chaque itération de la phase d'entraînement d'un modèle, un sous-ensemble de points de données d'entraînement est utilisé pour mettre à jour les paramètres du modèle. Le degré auquel cette mise à jour améliore la prédiction du modèle sur les données de validation donne une mesure d'utilité de ces points de données. Cette méthode est assez rapide et semble ne pas affecter considérablement le temps de calcul. Toutefois, elle utilise des approximations et n'échantillonne pas tous les sous-ensembles, ce qui peut affecter la précision de l'attribution. De plus, quoi que prometteuse, cette approche manque de validation empirique à grande échelle, et n'est donc pas utilisable en l'état. En outre, elle n'est pas utilisable pour chaque output généré par l'utilisateur, ce qui signifie qu'elle permet

essentiellement de mesurer la contribution d'un ensemble de données à la performance générale du modèle.

En conclusion, de nombreuses méthodes sont en cours de développement pour quantifier la contribution de différents jeux de données aux sorties générés par les modèles d'IA sur requêtes des utilisateurs, ou sur la performance globale du modèle. Sur le plan théorique, une méthode inspirée de la valeur de Shapley qui donne des résultats précis permettant une quantification juste apparaît idéale ; cependant, sa mise en pratique semble être rédhibitoire en termes de coûts (en temps et ressources de calculs) à grande échelle, en l'état actuel des connaissances. Elles pourraient être utilisées pour évaluer la performance globale des modèles sur un grand nombre de requêtes aléatoires ou représentatives, avec le risque que la contribution marginale de chaque donnée soit quasi-nulle. Elles ne pourraient pas être mises en place pour déterminer la contribution de chaque donnée à chaque sortie du modèle, quand bien-même procéderait-on par échantillonnage. Des approximations existent mais demeurent au stade de la preuve de concept. Une telle preuve a par exemple été fournie sur une base de données de faible taille (80 000 images comparé à des centaines de millions (DALL.E voire des milliards (Midjourney) d'images pour les modèles le plus utilisés(Bohacek & Farid, 2023)) et la solution est, à ce stade, loin d'être déployable à plus large échelle(J. T. Wang, Deng, et al., 2024)Wang et al., 2024). C'est pourquoi la recherche s'est orientée non sur un lien de causalité via l'utilité marginale mais vers une approximation plus déployable à l'échelle.

4.1.2 Etablir des liens de similarité entre la sortie du modèle et les données d'entraînement - Méthode « passive » corrélationnelle

Une seconde approche, alternative imparfaite à Shapley, réside dans la comparaison des caractéristiques des sorties avec des bases connues de données d'entrainement. Cette approche consiste à extraire certaines caractéristiques du contenu généré (comme, dans le cas de la musique, l'intensité, la tonalité, et la durée) et des données (e.g. d'entraînement), et de quantifier la similarité entre les caractéristiques de certaines données de la base d'entraînement et celles d'un contenu généré. Il n'est dans ce cas, pas besoin d'entraîner à nouveau un modèle déjà existant.

Des chercheurs de l'université d'Illinois Urbana-Champaign ont utilisé cette métrique pour vérifier que les données d'entraînement avec une plus forte valeur de Shapley étaient en effet plus similaires au son généré par le modèle d'IA (J. Deng et al., 2024). D'autres chercheurs(S.-Y. Wang et al., 2023) ont tout d'abord utilisé un modèle pré-entraîné pour l'affiner sur des images bien identifiées (chaque affinage est effectué sur une image « source »), sur la base desquelles ils ont généré des images synthétiques. De cette manière, les images synthétiques sont influencées par construction (par affinage) par l'image sur laquelle le modèle est affiné. Bien sûr, ces images synthétiques ne sont pas uniquement influencées par l'exemplaire sur lequel le modèle est affiné, mais cela suffit pour avoir une idée bruitée mais informative de l'image « source ».

L'idée est ensuite de tester différentes méthodes d'attribution, qui, si elles sont performantes, doivent être capables d'attribuer un score plus élevé à l'image source qu'à toute autre image du jeu d'entrainement. Les auteurs ont ensuite extrait des caractéristiques des images et mesuré la similarité entre les images du jeu d'entrainement et l'image générée, similarité convertie en probabilité que chaque image appartienne au jeu d'entrainement. Pour quantifier la capacité des modèles pré-entraînés (6 encodeurs couramment utilisés : DINO, CLIP, ViT, MoCo, SSCD, ALADIN) à retrouver le set d'entraînement, les chercheurs ont évalué la proportion d'éléments du jeu de données d'exemplaires ayant servi à la génération d'image dans le top-10 des images retrouvées (Recall@10). Les résultats obtenus sont largement au-dessus du niveau du hasard et varient en fonction de la méthode utilisée. De façon critique, les mêmes modèles pré-entraînés affinés sur les données d'attribution sont plus influencés par les exemplaires sources ayant servi à la génération d'images (cf. Encadré 4 Figure 3). Cependant, ils attribuent du crédit à beaucoup d'images qui

n'ont pas directement contribué à la genèse de la sortie du modèle, ce qui est problématique. Cette approche ne semble donc pas applicable en l'état même si des outils de démonstration rudimentaires existent, essentiellement à titre d'illustration (e.g. (*GenAI Attribution Simulator - a Hugging Face Space by TheFrenchDemos*, 2025; Lorphelin, 2024).

Cette seconde approche, moins coûteuse que Shapley semble fonctionner sur des images et sur un petit jeu de données comme le montrent les travaux menés actuellement par le pôle d'expertise de la régulation numérique (PEReN). Son coût computationnel n'est pas rédhibitoire, mais le résultat, en l'état actuel des recherches, serait d'une précision (très) limitée. Là encore, les études, récentes, sont au stade de la preuve de principe. Il n'est pas clair de savoir si cette approche pourrait être déployée à grande échelle et sur d'autres types de données. Une alternative serait d'échantillonner sur un nombre limité de requêtes par type d'IA (chatGPT, LLaMA, etc.) en supposant que l'échantillon est représentatif puis d'extrapoler à l'ensemble des requêtes. Enfin, rappelons que l'approche par similarité mathématique, notion détachée de celle de contrefaçon en PI, suppose de comparer un jeu de données entrantes de référence *connu* et un jeu de données sortantes *connu*.

Une autre piste consiste à prévenir les difficultés dans les prochains modèles en marquant les données d'entraînement d'un filigrane computationnel.

4.1.3 Marquage des données d'entrainement - Méthode proactive causale

Pour les modèles futurs, une solution pourrait être d'associer un filigrane à chaque image d'entraînement, et d'identifier ces marqueurs dans les images de sortie (Asnani et al., 2024). Par exemple, une méthode récente développée notamment par des chercheurs d'Adobe, ProMark, effectue l'attribution causale d'images synthétiques aux concepts prédéfinis présents dans les images d'entraînement.

Contrairement aux travaux précédents qui établissent une corrélation entre les images synthétiques et les données d'entraînement, cette méthode ne fait aucune supposition selon laquelle la similarité équivaut à une relation de causalité (cf. Encadré 4 Figure 4). ProMark associe des filigranes aux images d'entraînement et recherche ces filigranes dans les images générées, ce qui permet de démontrer directement la causalité plutôt que de simplement l'approximer ou l'impliquer. Le principe est simple : si un filigrane spécifique, unique à un concept d'entraînement (i.e. une image de pie, un ordinateur portable, etc.), peut être détecté dans une image générée, cela indique que le modèle génératif s'appuie sur ce concept lors du processus de génération.

Ainsi, ProMark repose sur deux étapes : le chiffrement des données d'entraînement via des filigranes et l'entraînement du modèle génératif avec des images filigranées. Pour filigraner les données d'entraînement, le jeu de données est d'abord divisé en N groupes, où chaque groupe correspond à un concept unique nécessitant une attribution. Ces concepts peuvent être sémantiques (par exemple, des objets, des scènes, des motifs ou des modèles d'images de stock) ou abstraits (comme des éléments stylistiques ou des informations de propriété). Chaque image d'entraînement d'un groupe est encodée avec un filigrane unique sans altérer de manière significative sa perceptibilité. Une fois les images d'entraînement filigranées, elles sont utilisées pour entraîner le modèle génératif. Au cours de l'apprentissage, le modèle apprend à générer des images à partir des images chiffrées. Idéalement, les images générées devraient contenir des traces des filigranes correspondant aux concepts dont elles sont issues.

Cette méthode permet bien de retrouver les images utilisées. Il s'agit, là encore, cependant, d'une preuve de concept qui n'est donc pas directement applicable en l'état. Toutefois, elle constitue une solution pour les futures données destinées à servir à l'entraînement.

Avant d'examiner la manière dont ces méthodologies pourraient, à l'avenir, être mobilisées pour quantifier des transferts de valeur avec les systèmes d'IA, il nous faut revenir sur les étapes actuelles de la valorisation des œuvres protégées.

4.2 Rémunération des œuvres protégées par le droit d'auteur : les étapes de la valorisation

Le droit d'auteur étant actuellement l'instrument au travers duquel s'effectue la rémunération des œuvres, il n'est pas inutile de rappeler les modalités de calcul de cette rémunération afin de mieux envisager les solutions dans le cas de l'IA. Sur le plan économique, la composante patrimoniale du droit d'auteur renvoie en effet à deux prérogatives distinctes :

- Le monopole d'autoriser ou d'interdire l'exploitation de l'œuvre sur un marché ;
- Une forme de rémunération en principe proportionnelle aux recettes d'exploitation de l'œuvre afin d'associer l'auteur au succès ; dans de nombreux cas des sommes de nature forfaitaire (avances non remboursables, minima garantis) versées avant toute recette sur le marché, complètent la rémunération proportionnelle.

Le paiement d'une rémunération implique plusieurs étapes de calcul :

- connaître l'assiette de rémunération (A),
- déterminer la part de cette assiette dévolue à l'amont de la filière (création- production) (P),
- répartir le montant perçu pour la création entre les diverses œuvres et ayants droit (R).

Chaque étape de ce calcul est confrontée à des problématiques diverses.

4.2.1 Une assiette de rémunération liée à l'activité de l'exploitant

En matière de propriété intellectuelle, deux grands principes permettent de définir l'assiette de rémunération :

- La rémunération doit être fonction du CA réalisé par l'exploitant de l'œuvre.
- La rémunération doit être en rapport avec l'exploitation des œuvres protégées.

Dans les cas les plus « purs » de la propriété intellectuelle, les deux approches se superposent : le CA de l'exploitant exprime, par le biais d'un prix payé par les usagers, une relation claire entre usages et œuvres. Les CA peuvent être de nature très différente (prix unitaire payé par le public, abonnements, revenus publicitaires, ressources publiques …).

Outre le CA, l'assiette de rémunération s'appuie systématiquement sur l'existence d'un lien fort entre l'activité et les pratiques des usagers concernant les œuvres concernées. Lorsque le CA provient de revenus publicitaires ou de la redevance des chaines de TV, des montants qui incluent des abattements sur l'assiette de rémunération sont négociés. Les CA qui n'ont aucun lien ou un lien trop lointain avec l'utilisation des œuvres ne sont pas retenus (ventes de confiserie dans les salles de cinéma par exemple).

L'existence d'un CA attribuable à certaines entreprises ou services, ne peut donc être prise comme base globale de rémunération lorsque le lien avec l'exploitation des contenus est faible ou très faible. Inversement, le fait que certaines entreprises choisissent de monétiser certaines activités et pas d'autres – proposées « gratuitement » à l'usager - est indifférent à l'importance des contenus pour les usagers dans les activités en question et **l'entreprise ne peut se prévaloir de l'absence de CA pour en déduire une assiette de rémunération nulle.**

4.2.2 Une part attribuée à l'amont modulée selon les habitudes professionnelles et les rapports de force entre les acteurs

Le poids accordé à l'amont de la filière (P) résulte surtout d'habitudes professionnelles et de négociations parfois anciennes pour les industries culturelles les plus installées. Dans l'édition par exemple, les auteurs perçoivent en moyenne 10% des recettes d'exploitation des librairies (Racine, 2020). Derrière ces moyennes se cache une répartition avec des taux très différents, notamment selon les segments de chaque filière et la notoriété des auteurs, négociés dans les contrats entre auteurs et éditeurs.

Pour les activités numériques plus récentes, cette part reflète tout autant une négociation et un rapport de force établi au sein de la filière. Pour la plateforme numérique de streaming Spotify, le chiffre d'affaires hors taxes de l'entreprise issu des revenus d'abonnement et des revenus publicitaires donne lieu après prélèvement de la part pour l'entreprise elle – même (environ 30%) à un montant destiné à la création et à la production (*Spotify Launches Revenue-Sharing Partner Program*, 2025) ; ce montant est lui – même réparti entre éditeurs, producteurs auteurs et artistes interprètes en fonction des contrats signés.

Le cabinet de conseil du Professeur Ernst Fehr s'est penché, quant à lui, sur la rémunération des médias (sur le marché suisse) dont les contenus sont proposés par Google search, qui devrait correspondre aux revenus publicitaires perdus par ces médias lorsque Google search détourne des utilisateurs de leur site web (Johann et al., 2023). Le cabinet tente d'établir la part du chiffre d'affaires qui devrait être allouée aux éditeurs de presse et procède, pour cela, en plusieurs étapes :

1) le montant du chiffre d'affaires sur le marché (suisse, 1 milliard CHF, provenant essentiellement des revenus publicitaires)
2) la part de marché relative au marché de la presse (le nombre de requêtes sur Google search relative à l'actualité, en l'occurrence environ 55%, soit 550 millions CHF)
3) la part de personnes qui n'auraient pas utilisé le service sans les contenus des ayants droit (le nombre de personnes qui n'aurait pas utilisé Google search si le résumé des articles de presse n'avait pas été fourni, soit 70% de 550 millions d'après une expérience, correspondant donc à 345 millions CHF)
4) le partage des revenus entre le moteur de recherche et les créateurs de médias, (ce qui d'après les pratiques correspond à 40% des 345 millions CHF, soit 154 millions CHF).

Au final, le montant (154 M), qui devrait selon cette méthode être dévolue aux éditeurs de presse correspond à environ 16% du chiffre d'affaires.

Les méthodes de calcul et le pourcentage obtenu sont cependant loin d'être incontestables ou généralisables à tous les marchés et à tous les services. Il s'agit pourtant de questions majeures. L'exemple du droit voisin des éditeurs de presse, instauré par le législateur européen puis transposé en droit français, est éclairant ; l'affirmation d'un principe, sans approfondir la délicate question des informations économiques nécessaires à la quantification de l'assiette de rémunération et de la part de cette assiette revenant aux éditeurs et celle revenant aux journalistes, montre combien cette disposition a conduit à de multiples difficultés d'implémentation.

De plus, en droit français, la rémunération proportionnelle au succès ne s'applique pas si la nature ou les conditions de l'exploitation rendent impossible l'application de ce type de rémunération, notamment lorsque la base de calcul de la participation proportionnelle ne peut être pratiquement connue ou déterminée. Dans ce cas c'est vers une rémunération forfaitaire (cf. supra) qu'il

convient de se tourner tout en déterminant, au préalable, les éléments de base sur lesquels s'appuyer pour fixer ce forfait.

Les rapports de force entre les acteurs jouent donc un rôle important dans le calcul du partage de la valeur (base sur laquelle partager puis pourcentages ou forfaits selon les cas) ; dès lors, la question qui peut être légitimement posée est une question de droit de la concurrence : dans quelle mesure, compte tenu d'éventuelles positions dominantes, le partage de valeur est-il négocié dans des conditions non discriminantes ? Veiller à ce que les négociations de partage de valeur aient lieu dans des conditions de concurrence loyale est donc un enjeu essentiel.

4.2.3 La répartition entre les œuvres et les ayants droit

La base forfaitaire accordée à l'amont que l'on retrouve dans de nombreuses situations, n'empêche pas, parfois ensuite, de tenter de rapprocher la répartition finale du succès estimé de chacun. Dans les cas les plus simples, la répartition entre les œuvres et entre les ayants droit de ces œuvres peut s'effectuer selon les données réelles, connues du succès. Dans d'autres cas, des sondages sont utilisés : un échantillon de 120 discothèques en France équipées d'un boitier permettant ainsi à la SPRE de connaître la diffusion de chaque morceau et d'établir la rémunération équitable associée pour chaque ayant droit (*La SPRE collecte la rémunération équitable pour les artistes-interprètes et les producteurs de phonogrammes*, n.d.; Lorphelin, 2024).

Dans le cas de Spotify, en 2023, la firme a généré 12.5 Mds de dollars de revenus et a reversé 9.5 Mds aux producteurs, aux éditeurs et aux auteurs compositeurs (*Loud and Clear by Spotify*, 2023). Le montant reversé à la filière musicale par la plateforme de streaming est ensuite réparti entre les différents acteurs de manière proportionnelle à la fréquence d'écoute des usagers (*Spotify Launches Revenue-Sharing Partner Program*, 2025). Mais les modalités de calcul de la répartition associée ont donné lieu à d'intenses débats entre l'approche « *market centric* » choisie par Spotify, « *user centric* » défendue par Deezer ou plus récemment l'approche contractuelle « *artist centric* ».

4.3 Quantifier les transferts de valeur dans le cas de l'IA

4.3.1 Lignes directrices

Quantifier les transferts de valeur nécessite donc de procéder en trois étapes complémentaires que nous détaillons ci-dessous.

1 - L'assiette des transferts (A)

Que la forme de partage soit de nature forfaitaire ou proportionnelle, le calcul préalable de l'assiette repose sur l'identification précise des entreprises et des activités au sein de ces entreprises en lien avec l'utilisation d'œuvres protégées et sur l'évaluation de la valorisation qui en résulte. Les données-œuvres s'intègrent dans un écosystème qui ne se limite pas à la phase d'entraînement ; les phases de déploiement, qui permettent notamment à un utilisateur, sur requête, de produire un résultat, sont sources de valorisations diverses. L'assiette de rémunération pourrait donc être élargie aux activités d'entreprises qui ne se contentent pas d'intervenir lors de la phase de pré- entraînement. Les chaînes de valeur étant encore en construction et les liens en cascade entre les services et applications nombreux (cf. partie 3), ce travail d'identification reste à approfondir. L'objectif est à la fois de cerner les services et applications concernés et d'éviter les doubles comptabilisations. La mission propose donc que, avec l'aide des services du Ministère de l'économie, une cartographie précise des lieux de création de valeur et des marchés pertinents sur lesquels asseoir le partage de valeur puisse être mise à l'agenda.

2 - La part dévolue à la création culturelle (P)

En matière de calcul de la part dévolue à l'amont de la filière (P) - qu'elle soit de nature proportionnelle ou forfaitaire - aucune règle claire, unique, sur la base d'un calcul économique défini, ne semble s'être imposée dans l'histoire des industries culturelles, en dehors de négociations, souvent complexes, entre les acteurs (voir supra). Dans le cas des systèmes d'IA, quelques points méritent à ce stade d'être soulignés. Tout d'abord, **la question du partage IA/culture doit être négociée pour des jeux de données (des « catalogues » d'œuvres) et non œuvre par œuvre.** D'autre part, la question de l'évaluation économique de la rémunération diffère de la question du principe juridique de la rémunération.

La transparence sur l'accès aux inputs en amont : socle insuffisant d'évaluation de la rémunération

Faire des obligations de transparence, prévues par *l'AI act* et de l'exercice de *l'opt-out*, prévu par la directive DAMUN, les bases d'évaluation de la rémunération apparaît comme largement insuffisant sur un plan économique. Les travaux juridiques conduits à l'échelle européenne sous l'égide de la Commission et du bureau de l'IA montrent que les exigences de transparence pourraient ne pas aller jusqu'au niveau de granularité attendu de certains titulaires de droits, ni concerner l'ensemble des acteurs de l'IA. Surtout, l'implémentation d'une transparence à la fois fine et respectant le secret des affaires, aboutirait certes à identifier la présence de sites ou de catalogues d'œuvres protégées au sein des inputs ; elle serait un des moyens d'appui pour l'indispensable exercice de droits au recours et à la preuve sur le plan juridique, ouvrirait la voie à des actions en justice et à une meilleure connaissance des partenaires avec lesquels négocier (sur les mécanismes alternatifs du droit à la preuve, cf. la partie juridique). Mais une fois la transparence acquise, rien ne serait clair sur la valeur économique de « l'emprunt » d'un jeu de données pour le modèle et sur la rémunération appropriée. Autrement dit, pour reprendre une métaphore culinaire, la transparence donne à voir la liste des ingrédients présents dans les placards de la cuisine mais ne dit pas dans quelles proportions chacun a été utilisé dans un plat donné – sans même parler de la recette - et par conséquent combien le cuisinier devrait payer pour obtenir ces ingrédients.

La destination des outputs en aval : socle indicatif complémentaire d'évaluation de la rémunération

Pour passer de la présence d'un jeu de données-œuvres à sa valorisation économique – au niveau « P » dans lequel nous nous situons ici - nous proposons de privilégier la destination d'un système d'IA en aval afin d'apprécier la valeur des données en amont. Il s'agit donc d'identifier au sein de la chaine de valeur des systèmes d'IA et de leurs applications, ceux dont l'activité a un lien direct avec l'utilisation des œuvres et ceux qui ont accès aux œuvres mais dont la destination du système d'IA est plus lointaine. Le niveau de rémunération P correspondant à l'exploitation de données – œuvres aux divers stades de développement et de déploiement des systèmes d'IA serait fondé sur une présomption économique [2] d'utilisation selon la destination du modèle.

La notion de « destination » est explicitement introduite dans l'*AI act* ; il s'agit de l'utilisation à laquelle un système d'IA est destiné par le fournisseur, y compris le contexte et les conditions spécifiques d'utilisation, tels qu'ils sont précisés dans les informations communiquées par le fournisseur dans la notice d'utilisation, les indications publicitaires ou de vente et les déclarations, ainsi que dans la documentation technique.

[2] On notera que la prise en compte d'une présomption économique liée à l'activité d'une organisation, aux seules fins d'évaluer le niveau de rémunération devant être accordée à la culture, est à la fois différente et complémentaire du débat sur l'instauration d'un éventuel mécanisme juridique de présomption, qui suppose des changements législatifs (cf. partie juridique).

La destination à laquelle nous faisons référence n'est pas celle des données entrantes mais celle, visible, des résultats produits par les modèles, les systèmes ou les applications. En effet, en matière d'IA, est parfois évoquée **l'idée de prix par grandes catégories de données voire, selon les phases du développement (pré-entraînement, fine tuning, RAG..) puisque la valeur de certaines données n'est pas la même selon les cas** (cf. partie 3). **Ces solutions mécaniques, nous semblent cependant ne pas devoir être retenues pour des raisons opérationnelles ; elles conduiraient sans aucun doute à des effets de bord et à des comportements opportunistes.** En l'absence de visibilité sur la destination des données entrantes – cas le plus courant, sauf lorsque les données concernées font l'objet d'un « marquage » systématique et généralisé (cf. supra 4.1.3) - il ne serait pas possible d'éviter, par exemple, qu'un acteur ait accès à bas prix à des données « d'entraînement », pour céder, bien plus cher, l'exploitation de ces mêmes données à des finalités de *fine tuning*.

La valorisation selon la destination des sorties ne correspond donc pas à l'idée, parfois avancée, de prix uniques par grandes catégories de données entrantes. La mission se prononce pour que les tarifs pratiqués, soient proposés par les titulaires de droits aux acteurs d'IA en fonction de catégories d'usages et d'usagers. Le principe de destination est d'ailleurs déjà habituellement pratiqué en matière de propriété intellectuelle pour moduler les niveaux de rémunération (Senftleben, 2024). Ainsi, dans l'exercice de la gestion collective musicale, la SACEM module les tarifs d'utilisation de son répertoire selon que le modèle d'affaires de l'exploitant repose de manière essentielle ou accessoire sur l'utilisation d'œuvres protégées : le coiffeur qui propose un fond sonore à sa clientèle, s'acquittera donc d'un bien plus faible pourcentage de son chiffre d'affaires que ne le fera une discothèque.

Pour l'alimentation des systèmes d'IA, il convient de distinguer, comme nous l'avons déjà noté, le fait déclencheur d'une rémunération (cf. partie juridique) et le niveau relatif de cette rémunération, auquel nous nous attachons ici. Le rôle de la mission n'est pas de rentrer dans +le calcul précis des tarifs de chacun dans tous les secteurs et pour tous les cas d'usage ; des lignes directrices, correspondant à trois grandes catégories de niveaux de rémunération, peuvent cependant schématiquement être identifiées, selon le critère de destination. Sur la base de ces catégories, un continuum de niveaux de tarification différents pourrait être établi par les opérateurs culturels eux-mêmes selon des licences annuelles, renégociables chaque année avec les opérateurs d'IA.

Destination n°1 : Niveaux de rémunération de base

La destination laisse présumer la mobilisation d'une vaste quantité de données indifférenciées dans laquelle le seul accès à des données protégées occupe une place accessoire pour la performance du modèle, parce-que d'autres données seraient substituables. Le modèle n'apprend pas sur des données particulières ou à des fins spécifiques.
Exemple : un modèle a mobilisé des textes littéraires lors de son entraînement ; une de ses applications consiste à répondre aux requêtes des clients d'une société d'assurance.

Destination n°2 : Niveaux de rémunération intermédiaires

La destination laisse présumer la mobilisation de données protégées non substituables en amont mais le modèle produit des outputs qui ne sont pas des quasi-œuvres susceptibles de remplacer les œuvres humaines.
Exemples : Une application fournit des outils de représentation tactile de morceaux musicaux pour les mal entendants ou des outils identifiant à la volée les œuvres écoutées ou encore des outils permettant aux producteurs de phonogrammes de gagner en productivité.

La valeur spécifique du résultat obtenu grâce aux inputs données-œuvres, conduit à un niveau de rémunération « intermédiaire » y compris lorsque ce résultat ne porte pas préjudice directement à un ayant droit précis ou n'apparaît pas « similaire » à un input. Des données- œuvres spécialisées et non substituables ont en effet été mobilisées pour aboutir à la performance globale du modèle.

Destination n°3 : Niveaux de rémunération supérieurs

La destination laisse présumer la mobilisation de données protégées non substituables en amont ET le modèle produit des quasi-œuvres synthétiques susceptibles de concurrencer les œuvres humaines.
Exemples : Un modèle génère automatiquement des images synthétiques d'illustration, susceptibles de remplacer les œuvres, ou le style (« à la manière de ») des illustrateurs humains.

3 - La répartition (R)

En matière de répartition, une méthode est celle du « *pay to train* » actuellement utilisée notamment dans le domaine audiovisuel ou de l'image. Les grandes banques d'images (Schutterstock, Getty images …) après avoir négocié avec des fournisseurs de modèles d'IA comme Open AI ou Google, répartissent en effet les montants perçus entre les acteurs selon la méthode de « pay to train » qui lie la rémunération au nombre d'œuvres que chacun détient dans le data set : un auteur détenant 200 photos dans le jeu de données recevra ainsi une rémunération double de celui détenant 100 photos. Shutterstock offre en moyenne et tous les 6 mois 0.0078 USD par image aux créateurs de contenus, avec une moyenne de 46 USD par portfolio (estimation effectuée sur un faible échantillon de 58 personnes) avec une base de données riche de 615 millions d'images, rémunérés en moyenne 0.0078 USD par unité, le montant reversé serait environ de 4.797 m USD, soit environ 2.2% du revenu (215.3 m USD en hiver 2023) et 15% du profit (32 m USD) de Shutterstock (Growcoot, 2023). Cette méthode de répartition, approximative, ne tient compte ici ni de la qualité des images ni de leur contribution marginale respective pour la requête d'un utilisateur donné. Dans l'audiovisuel, d'autres accords individuels passés avec des agrégateurs de contenus tels que la société américaine Calliope Networks valorisent les contenus sur la base de leur qualité, nature & durée (plus de 6 $ la minute pour des contenus exclusifs, avec majoration si qualité 4K ou 3D, vs 1 $ pour des formats courts), pour reverser ensuite les sommes collectées aux titulaires de droits avec lesquels la société a négocié des accords (la société est passée d'un catalogue de 17.000 heures de contenus audiovisuels en août 2024 à 35.000 heures début 2025 cf. https://calliopenetworks.ai/).

Ces méthodes de répartition ont l'avantage de la simplicité. Les modalités de répartition mobilisant de nouvelles méthodes de quantification pourraient permettre d'affiner en tenant compte de la contribution marginale de certaines données aux résultats produits.

4.3.2 Apport opérationnel des méthodes de quantification

Les différentes méthodes de quantification présentées dans la partie précédente, doivent, en effet, nous aider à approximer la contribution d'un jeu de données-œuvres et sa valorisation. La mise en œuvre de ces méthodes, quelles que soient leurs différences, a des limites qui ne permettent pas de les déployer en toutes occasions. La question économique posée devient donc celle de l'arbitrage, en termes de coûts de transaction, entre d'un côté le coût des méthodes de quantification et, d'un autre côté, les bénéfices espérés en termes de rémunérations.

Lorsque la quantité de données est telle que le gain marginal attendu pour un catalogue d'œuvres défini serait minime, potentiellement inférieur au coût de calcul, les méthodes de causalité et de similarité sont inopérantes. Tel sera le cas s'il s'agit d'évaluer la contribution d'un jeu de données aux résultats d'un modèle généraliste de fondation. Pour ces modèles

en effet, chaque jeu de données a rarement une influence décisive. Cela ne signifie pas que la valeur des données entrantes est nulle mais que les techniques de quantification actuelles ne permettent pas de déterminer cette valeur sans explosion des coûts. Concrètement, si le coût de l'implémentation d'une méthode d'attribution à chaque usage (ou presque si on échantillonne) du modèle égale ou excède les bénéfices, plus aucun bénéfice n'est à partager.

Dans les cas pour lesquels des catalogues d'œuvres bien identifiés ont été mis à la disposition des fournisseurs d'IA, **l'objectif est alors l'évaluation de la contribution relative de ces œuvres protégées aux modèles spécialisés. Dans ce cas, un nombre limité de données sont mobilisées par le modèle, rendant les solutions économiquement possibles,** au sens où le revenu marginal serait supposé conséquent.

Des chercheurs ont proposé d'utiliser des **méthodes d'attribution causales**, consistant à modifier le modèle en l'entraînant sur des jeux de données tronqués pour quantifier leur contribution à la production d'un output donné (e.g. « génère une image dans le style de … ») ou à la performance globale du modèle (e.g. le modèle est-il capable de générer des outputs correspondant aux requêtes des utilisateurs ?). Idéalement, un ensemble de sous-modèles est ré-entraîné pour calculer la valeur de Shapley et déterminer, à partir des contributions marginales des jeux de données, la répartition finale des rémunérations attendues par chacun des ayants droit (cf. Encadré 3). Dans le cas envisagé, les données d'entrée, connues, sont en nombre limité (pour que la méthode de Shapley puisse être déployée). Par ailleurs, dans l'exemple présenté (utilisant Shapley) les revenus réalisés sur les marchés doivent être également connus afin d'avoir une base de répartition. Des approximations de la méthode de Shapley pourront aussi être déployées (en utilisant l'échantillonnage de l'espace des sous-modèles).

Sur des jeux de données plus importants, **des méthodes d'approximation moins coûteuses que Shapley, par similarité** (entre la base de données d'entraînement et l'output), sont possibles, notamment sur des images, mais elles sont techniquement d'une moins grande précision ; en effet, la contribution d'une image clé peut être diluée (faux négatifs) ou, inversement, la contribution d'images qui n'ont peu ou pas été utilisées peut-être surestimée (faux positifs). En l'état actuel, cette technique n'est donc pas totalement opérationnelle.

Compte tenu de la jeunesse des recherches sur la quantification, la mission propose que des études supplémentaires puissent être menées rapidement sur des données d'entrée à déterminer et sur la base d'échantillonnage de requêtes, en collaboration avec les travaux actuellement menés par le Pôle d'expertise de la régulation numérique (PEReN)[3]. Ces expérimentations permettront de passer des preuves de concept à l'opérationnalité des méthodes sur des cas d'usage. Soulignons notamment que les modèles de génération de contenus d'images, d'audio, de textes ou de vidéos ne pourront être traités exactement de la même manière.

***

Le rappel des étapes habituelles de valorisation des œuvres protégées a mis en avant l'importance de penser les transferts de valeur, dans le cas de l'IA, avec des méthodologies différentes selon ce que l'on cherche à déterminer : l'assiette de ces transferts c'est-à-dire les lieux de création de valeur par les opérateurs de l'IA ; la part dévolue à la culture lors du partage entre opérateurs de l'IA et opérateurs culturels ; la répartition entre les œuvres et les ayants droit au sein des filières culturelles. Pour l'évaluation du Partage, au-delà du principe général de destination et de

[3] Le PEReN est un service interministériel dans le domaine des sciences des données, des algorithmes et de l'intelligence artificielle, qui intervient comme socle d'expertise technique mutualisé auprès des services de l'Etat et des autorités indépendantes en charge de la régulation des plateformes numériques.

présomption sur lequel il est possible de s'appuyer, les techniques de quantification sont opérantes dans des cas limités. Ces techniques seront sans doute les plus opérationnelles d'une part *au niveau de la Répartition* afin de faire en sorte que l'ensemble des créateurs en amont bénéficient du Partage de la valeur et d'autre part afin de *prouver* l'utilisation des oeuvres. Les deux tableaux suivants résument ces analyses.

Méthodologies d'évaluation des transferts de valeur

1 – Niveau A - Assiette
Objectif : Identifier la création de valeur des fournisseurs et déployeurs d'IA.
Méthode : Collaborer à une cartographie avec les services du Ministère de l'économie.

2 – Niveau P – Part dévolue à la culture
Objectif : Contribuer à déterminer différents niveaux de partage de valeur.
Méthodes :
- Mobiliser le critère de destination visible des activités qui se nourrissent de données protégées pour effectuer une gradation du partage.
- Approfondir les techniques de quantification pour prouver et évaluer la contribution relative d'un jeu de données-œuvres à la performance d'un modèle spécialisé et/ou aux réponses à des requêtes précises.

3- Niveau R – Répartition
Objectif : Répartir les montants P entre les différentes œuvres et ayants droit.
Méthode : Approfondir les techniques de quantification sur des cas d'usage.

Apports complémentaires du principe de destination des modèles et des techniques de quantification pour évaluer le partage (IA/culture) et la répartition (au sein de la filière culturelle)

<table>
<tr><th rowspan="2">Destination</th><th colspan="3">Méthode</th></tr>
<tr><th>CAUSALITÉ</th><th>SIMILARITÉ</th><th>MARQUAGE</th></tr>
<tr><td>Destination 1

Modèles généralistes

Partage : niveaux de rémunération de base</td><td colspan="2">Inopérant</td><td rowspan="3">Objectif envisagé : contribution d'un jeu de données à un résultat spécifique

Opérant uniquement pour les modèles à venir et si absence de risques de contournement

Opérationnalité possible : preuve d'utilisation pour le Partage et la Répartition</td></tr>
<tr><td>Destination 2

Modèles spécialisés culture - médias sans concurrence sur les outputs

Partage : niveaux de rémunération intermédiaires</td><td rowspan="2">Objectif envisagé : contribution d'un jeu de données à la performance globale du modèle/ à un résultat spécifique <u>avec un coût de calcul important</u>

Opérant sur des jeux de données en nombre limité

Opérationnalité possible : Partage et Répartition</td><td rowspan="2">Objectif envisagé : contribution d'un jeu de données à la performance globale du modèle/ à un résultat spécifique, <u>avec une précision limitée</u>

Opérant sur des jeux de données en nombre limité

Opérationnalité possible : Partage et Répartition</td></tr>
<tr><td>Destination 3

Modèles spécialisés culture - médias avec concurrence sur les outputs

Partage : niveaux de rémunération élevés</td></tr>
</table>

Le graphique suivant illustre à titre d'exemple, la manière dont pourraient s'opérer les transferts de valeur dans l'hypothèse d'une combinaison entre un dispositif de place marché (cf. partie 2), l'évaluation de la création de valeur des opérateurs d'IA (cf. partie 3) et la mobilisation du critère de destination pour partager la valeur entre opérateurs culturels et opérateurs d'IA (cf. partie 4).

Création de valeur des opérateurs d'IA

Facturation à leurs clients par les opérateurs des systèmes d'IA des différents services

Partage de valeur IA/Culture

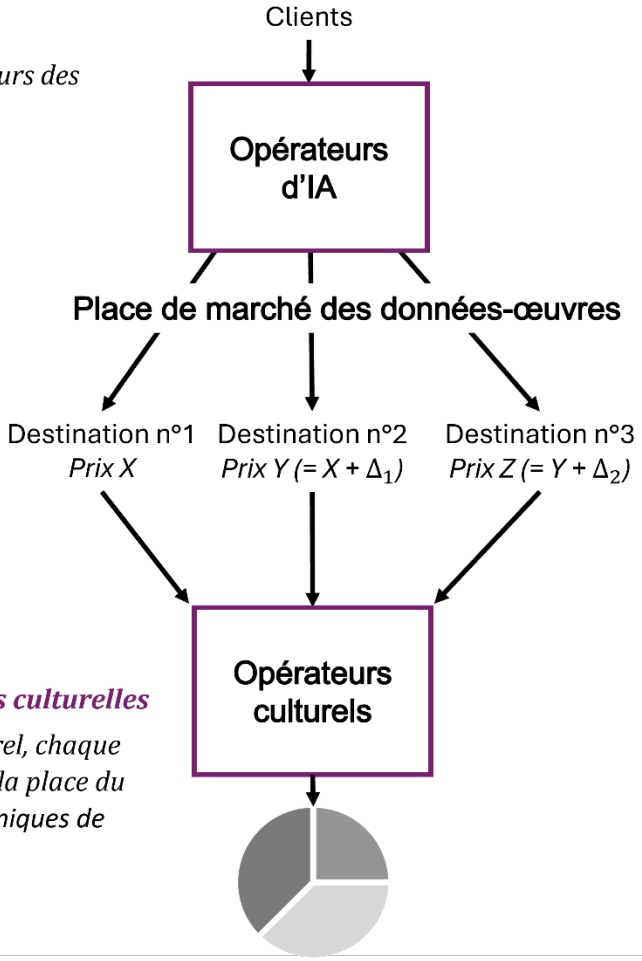
Place de marché regroupant les offres des divers opérateurs culturels (OGC, éditeurs, producteurs, …)

Détermination des prix

Prix des données déterminés par les opérateurs culturels selon la destination des systèmes d'IA dans le cadre de licences annuelles

Répartition de valeur au sein des filières culturelles

Répartition entre chaque opérateur culturel, chaque œuvre, chaque ayant droit représenté sur la place du marché, en s'appuyant sur différentes techniques de quantification



Recommandations issues du rapport économique

1 – Expliquer et médiatiser l'intérêt conjoint des opérateurs culturels et des opérateurs de l'IA, à investir dans un écosystème soutenable garantissant à la fois la présence des œuvres européennes dans les systèmes d'IA et la pérennité de leur financement.

2 – Mettre en place et/ou consolider des politiques d'accompagnement et de formation adaptées pour les métiers les plus directement impactés par l'essor de l'IA.

3 – Dans le cadre d'une concertation entre opérateurs culturels et opérateurs de l'IA, envisager l'opportunité et la faisabilité de la construction d'une place de marché, espace d'échanges structuré, permettant la contractualisation dans le respect des spécificités sectorielles.

4 - Dans le cadre d'une concertation entre opérateurs culturels et opérateurs de l'IA, explorer l'opportunité et la faisabilité des mécanismes de compensation et de transferts de valeur en complément de ceux fournis par la propriété intellectuelle.

5 – Réaliser avec les services du Ministère de l'économie, une cartographie précise de lieux de *création de valeur* et des marchés pertinents et suivre les circuits de valorisation dans la phase de déploiement, afin de fournir le socle du partage de valeur.

6 – Affiner l'opérationnalité du critère de présomption économique d'utilisation selon la destination des résultats produits par les modèles, systèmes ou applications qui se nourrissent de données protégées, afin d'effectuer une *gradation du partage* de valeur.

7 – En collaboration avec PEReN, approfondir sur des études de cas, l'opérationnalité des méthodes scientifiques de quantification pour prouver et/ou évaluer la contribution de certaines données – œuvres aux résultats produits et/ou à la performance globale des modèles spécialisés. Promouvoir auprès des opérateurs culturels et de l'IA les solutions jugées les plus pertinentes selon les cas d'espèce.

Organisations et personnalités ayant été entendues dans la cadre du rapport (volet juridique et économique)

Aday
Administration des Droits des Artistes et Musiciens Interprètes (ADAMI)
Adobe
AI Collaborative (Martin Tisné)
AI disclosure project (Ila, Stausss, Tim O'Reilly)
AIE (Associazione italiana editori)
Alliance de la presse d'information générale
Alliance français des industries du numérique (AFNUM)
Alt – Edic
Amazon
APIG
Association des développeurs et utilisateurs de logiciels libres pour les administrations et les collectivités territoriales (ADULLACT)
Association des Traducteurs Littéraires de France (ATLF)
Association Les voix
Autorité de la concurrence (ADLC)
Axel Springer
Bauer média
Bergeaud Antonin (HEC)
Bibliothèque nationale de France (BNF)
Bourreau Marc (Telecom Paris)
Brison Fabienne (Cabinet HOYng Rokh Monegier)
Cafeyn
Cairn (Thomas Parisot)
Centre français d'exploitation du droit de copie (CFC)
Centre national du cinéma et de l'image animée (CNC)
Combe Julien (Ecole Polytechnique)
Condé Nast
Controv3rse
Ekhoscènes
Emma ENPA
En chair et en os, collectif
Eurocinéma
Eviden
Fédération internationale de l'industrie phonographique (IFPI)
Federation of european publishers (FEP)
France Digitale
France télévisions
French Flair Entertainment
GEMA
Geste
Ginsburgh Jane (Columbia University)
Google
Groupement français des industries de l'information (GF2I)
Imatag
INRIA
Institut national de l'audiovisuel (INA)
L'Express
Lagardère
Les Echos – Le parisien
Ligue des auteurs professionnels

LinkUp
Ministère de l'économie, des finances et de l'industrie, Secrétariat d'Etat chargé de l'IA et du numérique, Cabinet
Ministère de l'économie, des finances et de l'industrie, Direction Générale du Trésor (DGT)
Ministère de l'économie, des finances et de l'industrie, Direction Générale des Entreprises (DGE)
Ministère de la culture, Cabinet
Ministère de la culture, Service des affaires juridiques et internationales
Miso.ai (Lucky Gunasekara)
Mistral
Netflix
Nouvelle République du Centre Ouest (NRCO)
Panneau Fabienne (Cabinet DLA Piper)
Perchet Vianney
Pôle d'Expertise de la Régulation Numérique (PEReN)
Prisma Media
Radio France
Rolling Stone magazine
Roux Steinkühler
Société civile des auteurs multimédia (SCAM)
Société de Perception et de Distribution des Droits des Artistes-Interprètes (SPEDIDAM)
Société des auteurs dans les arts graphiques et plastiques (ADAGP)
Société des auteurs des arts visuels et l'image fixe (SAIF)
Société des auteurs et compositeurs dramatiques (SACD)
Société des Auteurs, Compositeurs et Éditeurs de Musique (SACEM)
Société des gens de lettres (SGDL)
Société des Industries de la Presse et des Affiches (SIPA – Ouest France)
Société des Producteurs de Cinéma et de Télévision (PROCIREP)
Société Française des Intérêts des Auteurs de l'Écrit (SOFIA)
Syndicat de la presse indépendante d'information en ligne (SPIIL)
Syndicat des catalogues de film de patrimoine (SCFP)
Syndicat Français des artistes-interprètes (SFA-CGT)
Syndicat National de l'Édition (SNE)
Syndicat National des Auteurs et des Compositeurs (SNAC)
Syndicat national des éditeurs phonographiques (SNEP)
Syndicats National des Artistes Musiciens (SNAM-CGT)
TF1
Treppoz Edouard (Paris 1)
TrustMyContent
United Voice artists (UVA)

Annexes

Encadré 1 - Modèles de langage, modèles de diffusion et mesures de l'effondrement

L'effondrement correspond à la diminution de la qualité des données synthétiques lorsque les modèles de nouvelles générations sont entraînés sur la totalité ou sur une forte proportion de données synthétiques issues de la précédente génération de modèles. Elle a été mise en évidence dans de nombreux modèles d'IA, notamment les grands modèles de langage de type GPT et dans les modèles de diffusion pour la génération d'images. Quelle est la métrique utilisée pour ces deux types de modèles ? Comment reflète-t-elle l'effondrement ?

A) Modèles de langage et perplexité

Un modèle de langage attribue des probabilités à des séquences de symboles arbitraires de manière à ce que plus une séquence est susceptible d'exister dans cette langue, plus la probabilité attribuée est élevée. Un symbole peut être un caractère, un mot ou une sous-unité (par exemple, le mot « lire » peut être divisé en deux sous-unités : «li » et « re »). La plupart des modèles de langage estiment cette probabilité comme un produit de la probabilité de chaque symbole étant donné ses symboles précédents.

Soit une séquence $(w_1, w_2, \dots, w_n)$*, la probabilité de cette séquence est donnée par le produit suivant :*

$$P(w_1, w_2, \dots, w_n) = p(w_1)p(w_2|w_1)p(w_3|w_1, w_2) \dots p(w_n|w_1, w_2, \dots, w_{n-1})$$
$$= \prod_{i=1}^{n} p(w_i|w_1, \dots, w_{i-1})$$

(« $p(w_2|w_1)$ » se lit « probabilité d'occurrence de w_2 étant donnée l'occurrence de w_1 ». On calcule la probabilité de l'élément $w_2$sachant l'élément w_1.)

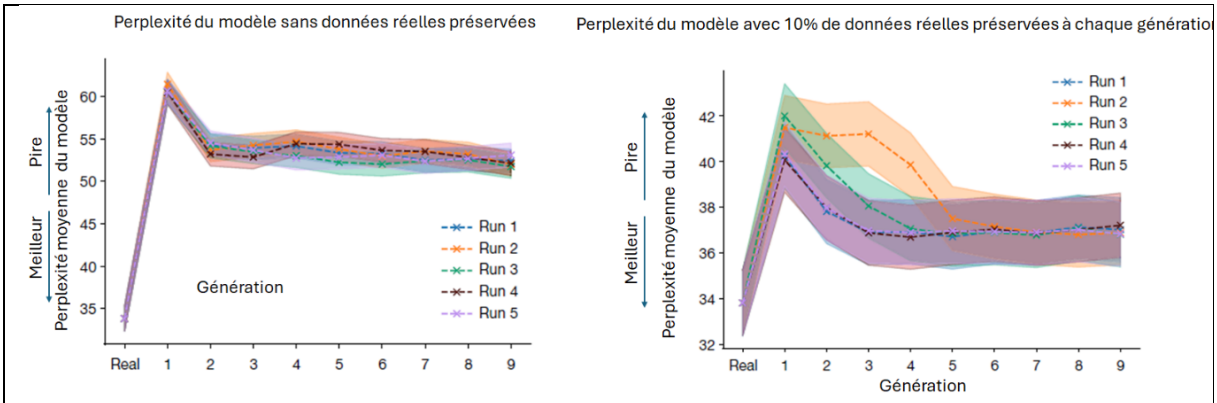
Autrement dit, la probabilité de la séquence « S = J'aime lire ce rapport » peut être calculée comme suit :
$P(S) = P(j') \times P(aime|j') \times P(lire|j'aime) \times P(ce|j'aime\ lire) \times P(rapport|j'aime\ lire\ ce)$

La perplexité est une mesure quantifiant à quel point une probabilité (e.g. générée par un modèle de langage) prédit un échantillon (e.g. le mot « rapport » après « J'aime lire ce »). La perplexité est une mesure d'incertitude relative à l'occurrence du prochain symbole. Mathématiquement, il s'écrit :

$$Perplexité\ (S) = exp\left\{-\frac{1}{n}\sum_{i}^{n} log\ p_\theta(w_i|w_{<i})\right\}$$

où $logP_\theta$ *est le logarithme de la vraisemblance d'obtenir le symbole* w_i *sachant les précédents symboles. Intuitivement, c'est une évaluation de la capacité du modèle à prédire les séquences d'un corpus.*

Les graphiques ci-dessous illustrent l'évolution de la perplexité quand un modèle de prochaine génération est entraîné sur des données synthétiques issues du modèle de la génération précédente exclusivement (à gauche) ou sur des données mixtes composées à 10% de données réelles et à 90% de données synthétiques du modèle de la génération précédente (à droite). Ces graphiques sont adaptés d'une des études présentées dans le rapport de (Shumailov et al., 2024).



Le terme « run » fait référence à une « expérience » : les auteurs ont effectué cinq expériences, c'est-à-dire qu'ils ont entraîné les générations de modèles sur la base de données synthétiques issues des précédentes générations, et cela cinq fois. L'objectif de cette approche est de s'assurer de la régularité statistique des observations et de la réplicabilité des résultats.

On observe bien que la perplexité du modèle entraîné sur les données réelles (« real ») est plus faible que la perplexité des modèles entraînés sur les données purement synthétiques ou mixtes (à partir de la génération 1 et jusqu'à 9).

B) Modèles de diffusion et distance de Fréchet (Frechet Inception Distance, FID)

Plusieurs types de modèles permettent de générer des images. Les modèles de diffusion consistent en l'induction de bruit (i.e. à « brouiller ») dans les images par étapes successives, et d'entrainer le modèle à débruiter les images jusqu'à retrouver l'image originale.

Il existe différentes mesures de la qualité de l'image générée par le modèle. L'une d'elles consiste à quantifier à quel point la distribution statistique des images générées par les modèles s'est éloignée de la distribution statistique des images présentes dans les données. Il s'agit de la distance de Wasserstein, qui mesure le travail minimum requis pour déplacer une densité de probabilité d'une distribution à une autre. En pratique, cette distance est approximée par la distance de Fréchet, appliquée aux caractéristiques extraites pour chaque image par un autre modèle déjà entraîné à cet effet (Inception, un réseau de neurones convolutifs) plutôt qu'à partir de chaque pixel. D'où le nom de la métrique (Frechet Inception Distance, FID).

L'idée est de ne pas travailler sur les pixels de l'image, mais d'extraire des caractéristiques de l'image, tels que les structures, les formes, les textures, des angles jusqu'aux objets. Un ensemble de caractéristiques est ainsi obtenu pour chaque image. Ces caractéristiques sont quantifiées et forment des distributions statistiques à l'échelle de la base de données, pour les données réelles d'une part, et pour les données synthétiques d'autre part. Les distributions peuvent être comparées entre ces deux types de données et peuvent être caractérisées par leur moyennes (notées μ_r, μ_s pour les données réelles et les données synthétiques, respectivement), et leur covariance (notées Σ_r, Σ_s pour les données réelles et les données synthétiques, respectivement). Mathématiquement :

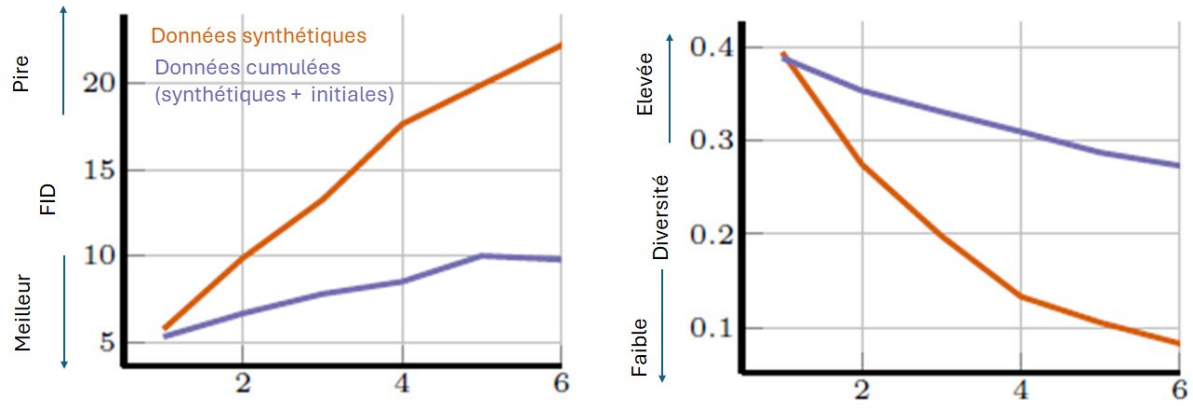
$$FID = \|\mu_r - \mu_s\|^2 + Tr\left(\Sigma_r + \Sigma_s - 2\sqrt{\Sigma_r \Sigma_s}\right)$$

où $\|\mu_r - \mu_s\|^2$ mesure la distance entre les moyennes des caractéristiques des données réelles et des données synthétiques ;
$Tr(.)$ représente la trace de la matrice, elle va évaluer la différence de variabilité entre les données générées et les données réelles. En effet, $Tr(\Sigma_r + \Sigma_s)$ reflète à quel point les deux distributions sont dispersées et $Tr\left(-2\sqrt{\Sigma_r \Sigma_s}\right)$ quantifie la similarité de la dispersion entre les deux distributions.

Un faible score de FID indique que les données générées sont proches des données réelles.

La diversité des données est mesurée par le rappel. La rappel (recall) estime la fraction d'échantillons dans une distribution de référence qui se trouvent dans le support de la distribution apprise par un modèle génératif (i.e. les valeurs réelles). Des scores de rappel élevés indiquent que le modèle génératif capture une grande partie des échantillons divers de la distribution de référence.

Les graphiques ci-dessous illustrent l'évolution de la FID (à gauche) quand un modèle de prochaine génération est entraîné sur des données synthétiques du modèle de la génération précédente exclusivement (en orange) ou sur des données mixtes composées des données réelles initiales plus toutes les données synthétiques cumulées à travers les générations. Le graphique de droite représente l'évolution de la diversité des images (le rappel). Ces graphiques sont adaptés d'une des études présentées dans le rapport Alemohammad et al., (2023).



On observe que la qualité (FID) du modèle entraîné sur les données synthétiques décroit (la FID augmente) au fil des générations de modèles entraînés sur des données synthétiques exclusivement (en orange) et, quoique dans une moindre mesure, pour l'accumulation de données synthétiques au côté des données réelles initiales (en bleu). La diversité suit le même pattern.

Encadré 2 – Les sources d'erreurs du processus d'effondrement

Le processus d'effondrement ***résulte de trois sources d'erreurs qui s'accumulent à travers les générations*** *et causent des déviations par rapport au modèle original* (Shumailov et al., 2024).

L'erreur d'approximation statistique initie le processus d'effondrement *du modèle éliminant les données extrêmes, affectant particulièrement les événements rares (i.e. en termes statistiques, en éliminant les queues de la distribution des données et d'autres détails statistiques fins de la distribution). Lorsqu'un modèle génère des données pour entraîner la prochaine génération de modèles, il peut ne pas inclure des événements rares ou de faible probabilité (comme des mots ou des enchaînements de mots inhabituels) qui étaient présents dans les données originales. Avec le temps, ces événements rares disparaissent, et la compréhension du modèle de la distribution des données d'origine se rétrécit, ce qui conduit à une dégradation des performances. Cette erreur provoque la disparition des queues de la distribution des données, ce qui signifie que le modèle se concentre davantage sur les événements courants et oublie les rares, menant ainsi à* ***son effondrement précoce.***

L'erreur d'expressivité fonctionnelle *affecte les premières étapes en limitant la capacité du modèle à capturer la complexité des données d'origine. Même si un modèle reçoit des données d'entraînement parfaites, il se peut qu'il n'ait pas une puissance expressive suffisante pour capturer avec précision la distribution sous-jacente réelle. En d'autres termes, le modèle peut être trop simple pour rendre compte de toutes les propriétés statistiques des données. Essayer d'approximer une distribution complexe (par exemple, un mélange de deux distributions gaussiennes) avec un modèle plus simple (par exemple, une seule gaussienne) entraînera des erreurs dans la manière dont le modèle représente cette distribution. Cette erreur survient principalement lors de la première génération du processus d'entraînement du modèle et peut entraîner des représentations incorrectes des données réelles, qui se propagent ensuite à travers les générations suivantes. Bien qu'elle ne puisse pas à elle seule expliquer entièrement l'effondrement du modèle, elle aggrave le problème dans les premières étapes. Tenter de contrer cet effet en augmentant la complexité du modèle peut avoir un effet pervers, celui d'expliquer le bruit dans les données et donc de générer des erreurs de généralisations. En d'autres termes, un modèle trop simple échoue à capturer toute la finesse des données, tandis qu'un modèle trop complexe risque de capturer du bruit stochastique (par exemple une régularité statistique comme un enchaînement particulier de phrases dans un texte propre à un jeu de données mais qui n'existe pas en général).*

L'erreur d'approximation fonctionnelle *résulte des limites dans les procédures d'apprentissage des modèles, qui peuvent induire des biais par exemple au cours de la descente de gradient ou selon le choix de l'objectif. Ces biais entrainent une déviation progressive par rapport aux données d'origine, en particulier dans les générations ultérieures. Même lorsque le modèle a une puissance expressive suffisante et des données d'entraînement abondantes, le processus d'apprentissage lui-même, par exemple, la manière dont le modèle optimise ses paramètres, ou le choix de ce qui est optimisé (e.g. la minimisation de l'erreur moyenne, la maximisation de la vraisemblance) introduit des biais qui peuvent entraîner une déviation supplémentaire par rapport à la distribution des données d'origine. Par exemple, le choix de ce qui est optimisé (la « fonction objectif ») joue un rôle critique. Si la fonction objectif utilisée pour entraîner le modèle minimise uniquement l'erreur moyenne, la procédure d'estimation du modèle peut ignorer des aspects importants, comme la préservation des événements rares (les "queues" de la distribution). Ces biais s'accumulent au fil des générations, à mesure que les modèles sont continuellement ajustés sur des données générées par des modèles antérieurs, ce qui entraîne une divergence par rapport à la distribution d'origine et* ***un effondrement tardif du modèle.*** *Au fur et à mesure que les modèles échouent progressivement à capturer la distribution des données d'origine, ils s'effondrent et produisent des sorties dégénérées ou simplifiées avec une faible variance.*

Encadré 3 : La méthode de la « Shapley value » appliquée à des jeux de données culturelles protégées

L'article de Wang et al 2024 mobilise la valeur de Shapley pour calculer dans quelle mesure les données de chaque titulaire de droits d'auteur ont contribué au succès d'un modèle d'IA génératif dans la création de contenu spécifique. Sur la base de cet article, nous proposons d'illustrer comment la valeur de Shapley pourrait être calculée dans un exemple simple, imaginé pour l'occasion.

Imaginons un modèle entraîné sur les données de plusieurs titulaires de droits d'auteur, et que l'on veuille déterminer dans quelle mesure les données de chacun ont contribué à la capacité du modèle à générer un contenu spécifique. La valeur de Shapley aide à répondre à cette question en prenant en compte tous les sous-ensembles possibles et en déterminant dans quelle mesure les données de chaque titulaire augmentent l'utilité du modèle lorsqu'elles sont ajoutées à diverses combinaisons de sous-ensembles de données.

Considérons le processus étape par étape avec un exemple de trois titulaires de droits d'auteur, A, B et C, chacun ayant contribué des données pour entraîner le modèle. On peut générer une image avec une IA comme DALL.E puis déterminer dans quelle mesure les données de chaque propriétaire ont contribué à cette œuvre. Supposons que :

- **A** possède un ensemble de données de paysages.
- **B** possède un ensemble de données de portraits.
- **C** possède un ensemble de données d'art abstrait.

Dans l'article de Wang et al. 2024 l'utilité est une mesure de la performance du modèle pour générer un résultat spécifique (par exemple, une œuvre d'art spécifique). L'utilité peut être vue comme la probabilité que le modèle génère la même œuvre d'art en utilisant les données d'un sous-ensemble donné de propriétaires.

Supposons que l'on mesure l'utilité des différents sous-ensembles comme suit :

- Utilité du sous-ensemble {A, B, C} (l'ensemble complet) : 100 (c'est l'utilité du modèle entièrement entraîné).
- Utilité du sous-ensemble {A, B} : 80 (A et B ensemble génèrent bien du contenu, mais pas aussi bien que le modèle complet).
- Utilité du sous-ensemble {A, C} : 80.
- Utilité du sous-ensemble {B, C} : 60.
- Utilité du sous-ensemble {A} : 50.
- Utilité du sous-ensemble {B} : 40.
- Utilité du sous-ensemble {C} : 30.
- Utilité de l'ensemble vide {} : 0 (pas de données, pas de modèle).

L'élément clé pour calculer la valeur de Shapley est d'examiner la contribution marginale de chaque propriétaire de données aux différents sous-ensembles. La contribution marginale est la mesure dans laquelle l'ajout des données d'un propriétaire particulier améliore l'utilité du modèle.

Par exemple :

- Ajouter **A** à l'ensemble vide {A} fait passer l'utilité de 0 à 50, donc la contribution marginale de A est **50** dans ce cas.
- Ajouter **B** à {A} fait passer l'utilité de 50 à 80, donc la contribution marginale de B est **30**.
- Ajouter **C** à {A, B} fait passer l'utilité de 80 à 100, donc la contribution marginale de C est **20**.

La valeur de Shapley est calculée en faisant la moyenne de la contribution marginale de chaque propriétaire de droits d'auteur sur toutes les façons possibles de combiner les données des propriétaires. La formule est :

$$\phi i = \frac{1}{n}\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!\,(n-k)!}\sum_{S\subseteq N\{i\}}[v(S\cup\{i\}-v(S)]$$

Où :

- n est le nombre total de titulaires de droits d'auteur.
- S représente un sous-ensemble de titulaires de droits d'auteur.
- v(S) est l'utilité du sous-ensemble S.
- La somme est effectuée sur tous les sous-ensembles de propriétaires qui n'incluent pas i.

La somme de gauche indique qu'il faut sommer à travers toutes les combinaisons possibles d'ordre de contribution de l'ayant droit i.

La somme de droite correspond à la contribution marginale de l'ayant droit i (c'est la différence de l'utilité des données avec i – utilité des données sans i).

Maintenant, calculons la valeur de Shapley pour chaque titulaire (A, B, C) en faisant la moyenne de leurs contributions marginales sur tous les sous-ensembles.

Par exemple :

- Pour **A** :
  - Contribution de A lorsqu'il est ajouté à {} : 50 (l'utilité passe de 0 à 50).
  - Contribution de A lorsqu'il est ajouté à {B} : 30 (de 40 à 80).
  - Contribution de A lorsqu'il est ajouté à {C} : 50 (de 30 à 80).
  - Contribution de A lorsqu'il est ajouté à {BC} : 40
  - Contribution marginale moyenne de A : $\frac{50+30+50+40}{4} = 42.5$

- Pour **B** :
  - Contribution de B lorsqu'il est ajouté à {} : 40.
  - Contribution de B lorsqu'il est ajouté à {A} : 30.
  - Contribution de B lorsqu'il est ajouté à {C} : 30.
  - Contribution de A lorsqu'il est ajouté à {AC} : 20.
  - Contribution marginale moyenne de B : $\frac{40+30+30+30}{4} = 32.5$

- Pour **C** :
  - Contribution de C lorsqu'il est ajouté à {} : 30.
  - Contribution de C lorsqu'il est ajouté à {A} : 30.
  - Contribution de C lorsqu'il est ajouté à {B} : 20.
  - Contribution de C lorsqu'il est ajouté à {A,B} : 20.
  - Contribution marginale moyenne de C : $\frac{30+30+20+20}{4} = 25$

Ainsi, les valeurs de Shapley seraient :

- $\Phi(A) = 42.5$
- $\Phi(B) = 32.5$
- $\Phi(C) = 25$

En fonction de ces valeurs de Shapley, les titulaires de droits d'auteur recevraient une rémunération proportionnelle à leurs contributions. Par exemple, si l'œuvre générée par le modèle rapporte 100 $, les gains pourraient être répartis comme suit :

- **A** reçoit environ 42.5 $.
- **B** reçoit environ 32.5 $.
- **C** reçoit environ 25 $.

Encadré 4. Les différentes méthodes d'estimation de la contribution de données d'entraînement à la sortie de l'IA générative.

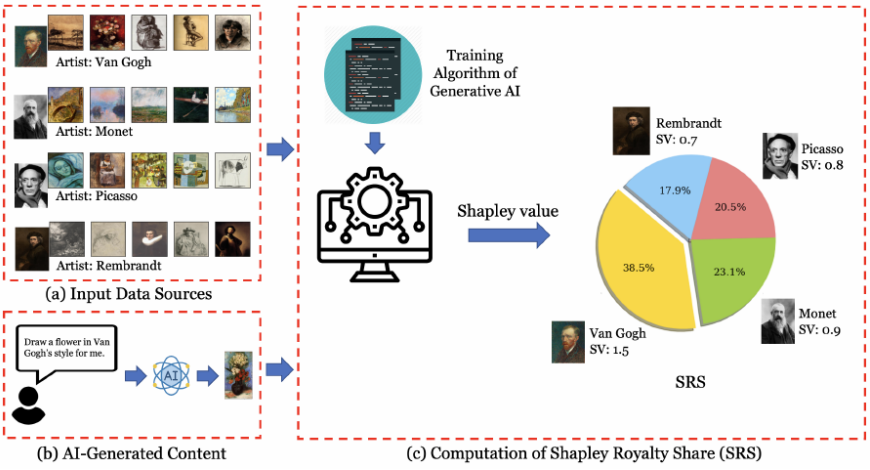


Figure 1. Vue d'ensemble sur le calcul de la valeur de Shapley (approche causale)

Nous reprenons ici les graphiques de (J. T. Wang, Deng, et al. 2024) aux seules fins d'illustration méthodologique. Les peintres cités dans cet article sont en effet dans le domaine public, donc non concernés par la question de partage de valeur, objet de cette mission.

a : supposons que quatre jeux de données soient utilisées pour entraîner un modèle d'IA
b : un utilisateur utilise une interface pour soumettre une requête au modèle d'IA générative : « dessine une fleur dans le style de Van Gogh ».
c : La valeur de Shapley peut être calculée de façon à déterminer la part due à chaque artiste dans la sortie générée. Le modèle est d'abord entraîné avec les données de Van Gogh, puis affiné sur celles de Monet, puis sur celles de Picasso, puis celles de Rembrandt, et cela dans divers ordres pour estimer la contribution marginale de chaque auteur, pour une sortie donnée. Cette méthode à un coût prohibitif et nécessite des approximations.

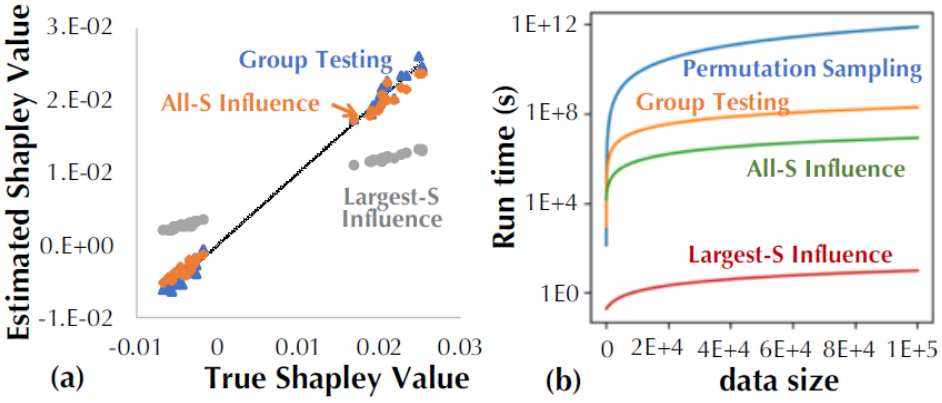


Figure 2. Comparaison de différentes méthodes d'approximation de la valeur de Shapley (approche causale)
D'après (Jia et al., 2019).
« Permutation sampling » correspond à l'entraînement de modèle sur une partie aléatoire des sous-ensemble de données ; « Group testing » correspond à l'entraînement des modèles sur plusieurs sous-ensembles de données suivant un échantillonnage « intelligent » basé sur la théorie des groupes ; All-S et largest-S Influence correspondent à l'utilisation de la méthode dite « Influence function » consistant à étudier l'importance d'un point de données en le surpondérant dans les modèles, entraînés sur toutes les données (largest-S) ou sur des sous-ensemble (all-S).

La Figure 2A montre l'estimation de la valeur de Shapley par la méthode approximative en fonction de la vraie valeur de Shapley. Si l'approximation est précise, on s'attend à ce que chacun des points soit aligné avec la ligne noire. C'est le cas de la méthode « group testing » et « all-s influence », mais pas de « largest-s influence.

La Figure 2B montre le temps de calcul pour chaque méthode (en échelle logarithmique), en fonction de la taille du jeu de données. Au total sur les deux figures, on constate que les méthodes rapides sont peu précises, et inversement.

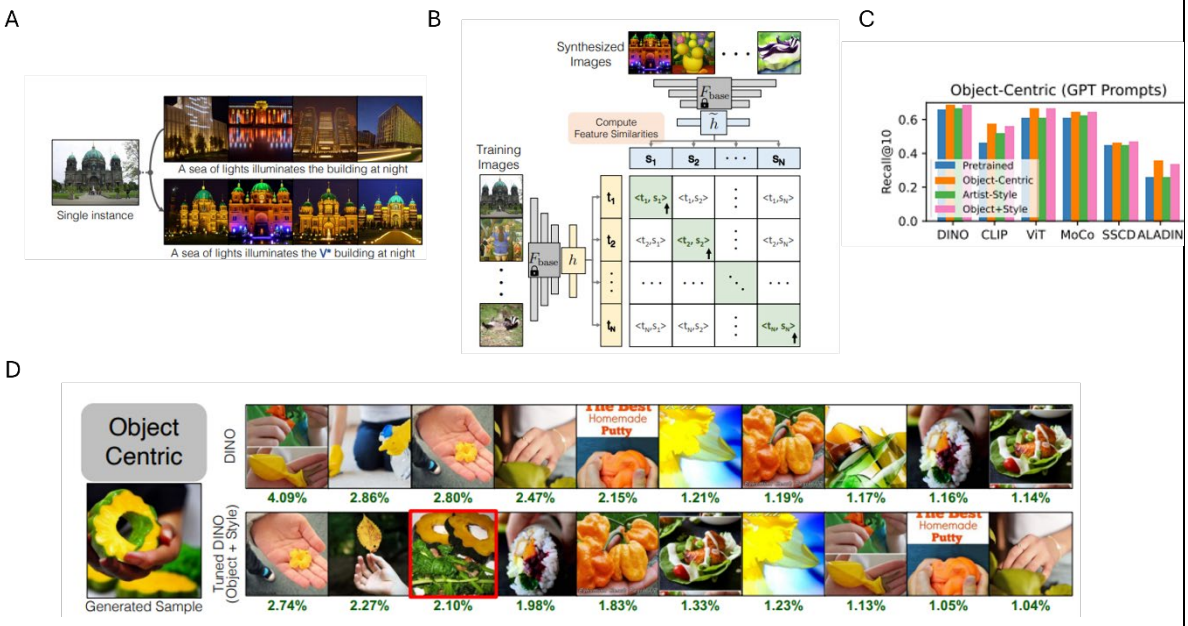


Figure 3. Exemple d'utilisation des caractéristiques de l'image pour calculer la similarité avec la sortie du modèle (approche corrélationnelle) D'après (S.-Y. Wang et al., 2023).

A : Illustration de la phase d'affinage sur une image source, en l'occurrence un bâtiment précis. Demander au modèle d'IA des images de bâtiments éclairés produit des bâtiments éclairés variés, tandis que demander de générer des bâtiments sur la base de l'image source génère différentes versions éclairées de ce bâtiment source.
B : Calcul de la similarité entre les images générées (les colonnes du tableau) en fonction de l'image source (les ligne du tableau). On s'attend à de forte similarité sur les diagonales du tableau, c'est-à-dire que les images générées doivent ressembler davantage à l'image source sur lequel le modèle est affiné qu'à tout autre image source.
C : Pour vérifier cette prédiction, les auteurs calculent la fréquence d'images sources dans le top 10 des images du jeu de données d'entrainement + affinage (Recall@10), pour chaque grand type de modèle (DINO, CLIP, etc.). La barre bleue concerne les performances du modèle préentraîné, la barre orange est celle qui nous intéresse dans cet exemple : elle doit être supérieure à la barre bleue pour que la méthode d'attribution puisse être considérée comme correcte.
D : Dans cet exemple, on peut voir que l'image source, encadrée en rouge, est bien dans le top 10 des images le plus similaires à l'image générée. On voit aussi que d'autres images, pourtant non utilisées, sont similaires et seraient donc créditées d'une rémunération (à tort).

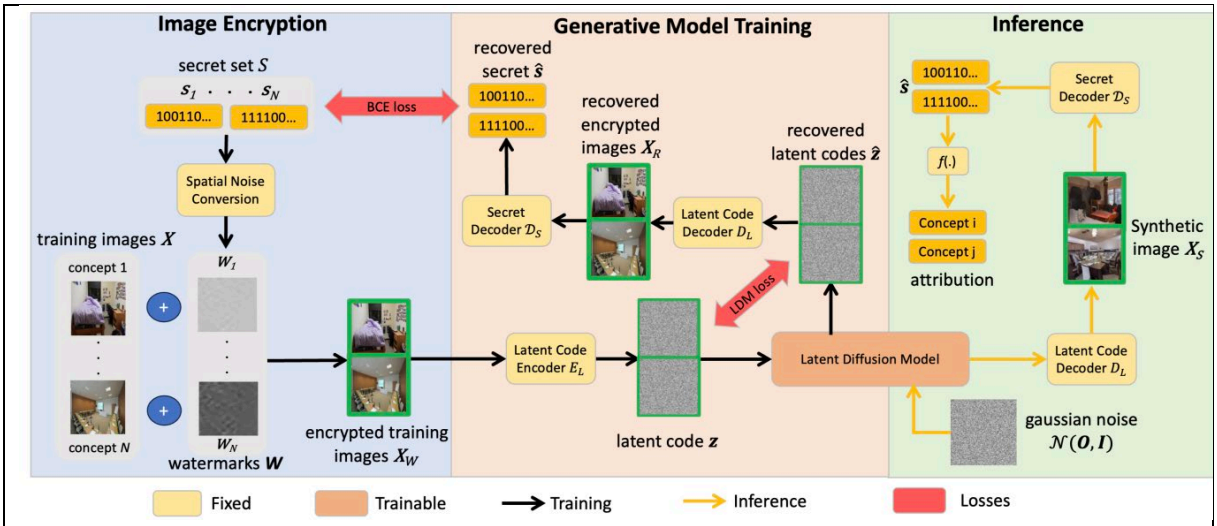


Figure 4. La méthode du marquage par filigrane (approche causale)
D'après (Asnani et al., 2024).
Dans la première phase, un filigrane « invisible » à l'œil nu est associé à chaque image. Dans la seconde phase, le modèle est entraîné sur les données d'entraînement marquées ; le marquage est retrouvé. Dans la phase d'inférence, c'est-à-dire de génération d'une image de sortie sur la demande d'un utilisateur, il est possible de décoder les images qui ont contribué à la genèse des images synthétiques.

Bibliographie

A three-step design pattern for specializing LLMs. (n.d.). Google Cloud Blog. Retrieved 12 March 2025, from https://cloud.google.com/blog/products/ai-machine-learning/three-step-design-pattern-for-specializing-llms

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2022). Artificial Intelligence and Jobs: Evidence from Online Vacancies. *Journal of Labor Economics*, *40*(S1), S293–S340. https://doi.org/10.1086/718327

Aghion, P., & Bouverot. (2024). *IA : notre ambition pour la France*. Comission de l'intellignece artificielle.

AI Foundation Models: Update paper. (2024, April 16). GOV.UK. https://www.gov.uk/government/publications/ai-foundation-models-update-paper

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A., & Baraniuk, R. G. (2023). *Self-Consuming Generative Models Go MAD* (arXiv:2307.01850). arXiv. https://doi.org/10.48550/arXiv.2307.01850

Asnani, V., Collomosse, J., Bui, T., Liu, X., & Agarwal, S. (2024). ProMark: Proactive Diffusion Watermarking for Causal Attribution. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10802–10811. https://openaccess.thecvf.com/content/CVPR2024/html/Asnani_ProMark_Proactive_Diffusion_Watermarking_for_Causal_Attribution_CVPR_2024_paper.html

Audiens data lab report. (n.d.). IA: mesure de l'évolution des métiers du doublage. Retrieved 10 November 2024, from https://app.powerbi.com/view?r=eyJrIjoiZTA1NGU0OTctMjRjNC00Y2VlLWExMDktNzk0YTRkMjg3NGVjIiwidCI6ImQwMGVkZjI4LTUyOTUtNGJmZS05YTU5LTM1NjA4NWMzZGU3YyJ9

Avis 24-A-05 du 28 juin 2024. (2024, June 28). Autorité de la concurrence. https://www.autoritedelaconcurrence.fr/fr/avis/relatif-au-fonctionnement-concurrentiel-du-secteur-de-lintelligence-artificielle-generative

Benabou, V.-L. (2018, November 14). *Mission du CSPLA sur les conséquences pour la propriété littéraire et artistique de l'avènement des notions de données et de contenus numériques*. https://www.culture.gouv.fr/nous-connaitre/organisation-du-ministere/Conseil-superieur-de-la-propriete-litteraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA/Mission-du-CSPLA-sur-les-consequences-pour-la-propriete-litteraire-et-artistique-de-l-avenement-des-notions-de-donnees-et-de-contenus-numeriques

Benabou, V.-L. (2023, December 11). Du test en trois étapes au domaine public payant- Quelques idées pour mieux associer les titulaires de droit à la production des Intelligences Artificielles génératives dans le champ de la création intellectuelle. *Le 'chat' et la souris*. https://vlbenabou.blog/2023/12/11/168/

Bensamoun, A., & Farchy, J. (2020). *Intelligence artificielle et culture* [Rapport du CSPLA]. Ministère de la culture.

Bertrand, Q., Bose, A. J., Duplessis, A., Jiralerspong, M., & Gidel, G. (2024). *On the Stability of Iterative Retraining of Generative Models on their own Data* (arXiv:2310.00429). arXiv. https://doi.org/10.48550/arXiv.2310.00429

Besson, L., Dozias, A., Faivre, C., Gallezot, C., Gouy-Waz, J., & Vidalenc, B. (2024). *Les enjeux économiques de l'intelligence artificielle* (341; Trésor-Eco). Direction Générale du Trésor. https://www.tresor.economie.gouv.fr/Articles/2024/04/02/les-enjeux-economiques-de-l-intelligence-artificielle

Bianchi, F., Kalluri, P., Durmus, E., Ladhak, F., Cheng, M., Nozza, D., Hashimoto, T., Jurafsky, D., Zou, J., & Caliskan, A. (2023). Easily Accessible Text-to-Image Generation Amplifies Demographic Stereotypes at Large Scale. *2023 ACM Conference on Fairness, Accountability, and Transparency*, 1493–1504. https://doi.org/10.1145/3593013.3594095

Bohacek, M., & Farid, H. (2023). *Nepotistically Trained Generative-AI Models Collapse* (arXiv:2311.12202). arXiv. http://arxiv.org/abs/2311.12202

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, *29*. https://proceedings.neurips.cc/paper_files/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

Business Insights and Conditions Survey Team. (2024). *Business insights and impact on the UK economy—Office for National Statistics*. Office for National Statistics. https://www.ons.gov.uk/businessindustryandtrade/business/businessservices/bulletins/businessinsightsandimpactontheukeconomy/3october2024

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Cantor, M. (2023). Nearly 50 news websites are 'AI-generated', a study says. Would I be able to tell. *The Guardian, May*.

Cazzaniga, M., Jaumotte, M. F., Li, L., Melina, M. G., Panton, A. J., Pizzinelli, C., Rockall, E. J., & Tavares, M. M. M. (2024). *Gen-AI: Artificial intelligence and the future of work*. International Monetary Fund. https://books.google.com/books?hl=en&lr=&id=YLXuEAAAQBAJ&oi=fnd&pg=PA2&dq=Gen-AI:+Artificial+Intelligence+and+the+Future+of+Work&ots=OOSfbJ70wz&sig=X5E0FQ-27qQLdeKPYMME4nh6Cek

Chin-Rothmann, C. (2023). *Navigating the Risks of Artificial Intelligence on the Digital News Landscape*. https://www.csis.org/analysis/navigating-risks-artificial-intelligence-digital-news-landscape

Cohen, C. (2024, May 26). *«L'IA a volé mon travail du jour au lendemain»: Ces traducteurs et doubleurs déjà remplacés dans le monde de l'édition*. Le Figaro. https://www.lefigaro.fr/medias/l-ia-a-vole-mon-travail-du-jour-au-lendemain-ces-traducteurs-et-doubleurs-que-l-ia-a-deja-remplaces-dans-le-monde-de-l-edition-20240526

Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296–299). Auerbach Publications. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003278290-44/amazon-scraps-secret-ai-recruiting-tool-showed-bias-women-jeffrey-dastin

Deng, C., Feng, S., Wang, H., Zhang, X., Jin, P., Feng, Y., Zeng, Q., Chen, Y., & Lin, Y. (2022). OpenFWI: Large-scale multi-structural benchmark datasets for full waveform inversion. *Advances in Neural Information Processing Systems*, *35*, 6007–6020.

Deng, J., Zhang, S., & Ma, J. (2024). *Computational Copyright: Towards A Royalty Model for Music Generative AI* (arXiv:2312.06646). arXiv. https://doi.org/10.48550/arXiv.2312.06646

Directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE (Texte présentant de l'intérêt pour l'EEE.)—Légifrance. (n.d.). Retrieved 22 November 2024, from https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038481211

Distilling step-by-step: Outperforming larger language models with less training. (n.d.). Retrieved 12 March 2025, from https://research.google/blog/distilling-step-by-step-outperforming-larger-language-models-with-less-training-data-and-smaller-model-sizes/

Dohmatob, E., Feng, Y., Yang, P., Charton, F., & Kempe, J. (2024). *A Tale of Tails: Model Collapse as a Change of Scaling Laws* (arXiv:2402.07043). arXiv. https://doi.org/10.48550/arXiv.2402.07043

Elias, G. (2024, June 11). *Profil de l'entreprise Runway ML - AI&YOU #44*. Skim AI. https://skimai.com/fr/aiyou-44-notre-profil-dentreprise-de-la-piste-ml/

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). *GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models* (arXiv:2303.10130). arXiv. http://arxiv.org/abs/2303.10130

European Commission. (2021). *EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act*. European Commission. https://artificialintelligenceact.eu/

Gault, M. (2023). AI spam is already flooding the internet and it has an obvious tell. *VICE, April*.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2022). *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness* (arXiv:1811.12231). arXiv. https://doi.org/10.48550/arXiv.1811.12231

GenAI Attribution Simulator—A Hugging Face Space by TheFrenchDemos. (2025). https://huggingface.co/spaces/TheFrenchDemos/genai-reward

Gerstgrasser, M., Schaeffer, R., Dey, A., Rafailov, R., Sleight, H., Hughes, J., Korbak, T., Agrawal, R., Pai, D., Gromov, A., Roberts, D. A., Yang, D., Donoho, D. L., & Koyejo, S. (2024). *Is Model Collapse Inevitable? Breaking the Curse of Recursion by Accumulating Real and Synthetic Data* (arXiv:2404.01413). arXiv. https://doi.org/10.48550/arXiv.2404.01413

Glickman, M., & Sharot, T. (2024). AI-induced hyper-learning in humans. *Current Opinion in Psychology*, *60*, 101900. https://doi.org/10.1016/j.copsyc.2024.101900

Growcoot, M. (2023, July 12). *Shutterstock May Have Paid Out Over $4 Million From its AI Contributor Fund*. PetaPixel. https://petapixel.com/2023/07/12/shutterstock-may-have-paid-out-over-4-million-from-its-ai-contributor-fund/

Hammoudeh, Z., & Lowd, D. (2024). Training data influence analysis and estimation: A survey. *Machine Learning*, *113*(5), 2351–2403. https://doi.org/10.1007/s10994-023-06495-7

Hoppner, T., & Streatfeild, L. (2023). ChatGPT, Bard & Co.: An introduction to AI for competition and regulatory lawyers. *An Introduction to AI for Competition and Regulatory Lawyers (February 23, 2023)*, *9*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371681

House of Lords Communications and Digital Select Committee inquiry. (2023, December 5). *OpenAI—written evidence (LLM0113)*. https://committees.parliament.uk/writtenevidence/126981/pdf/

Hughes, A. (2023, December 12). Phi-2: The surprising power of small language models. *Microsoft Research*. https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/

Hughes, A. (2024, March 5). Orca-Math: Demonstrating the potential of SLMs with model specialization. *Microsoft Research*. https://www.microsoft.com/en-us/research/blog/orca-math-demonstrating-the-potential-of-slms-with-model-specialization/

IA générative et droit d'auteur: Quelle place pour les données européennes protégées à l'ère de l'IA ? (2024). France Digitale. https://francedigitale.org/publications/ia-generative-et-droit-dauteur

Inquiets de l'utilisation de l'IA, les acteurs et doubleurs de jeux vidéo vont faire grève en Californie. (2024, July 26). Le Figaro. https://www.lefigaro.fr/medias/inquiets-de-l-utilisation-de-l-ia-les-acteurs-et-doubleurs-de-jeux-video-vont-faire-greve-en-californie-20240726

Jia, R., Dao, D., Wang, B., Hubis, F. A., Hynes, N., Gürel, N. M., Li, B., Zhang, C., Song, D., & Spanos, C. J. (2019). Towards efficient data valuation based on the shapley value. *The 22nd International Conference on Artificial Intelligence and Statistics*, 1167–1176. https://proceedings.mlr.press/v89/jia19a.html

Johann, A., Drazilova, M., Treweller, S., Möhlen, J., Brusoni, S., & Fehr, E. (2023). The Value of Journalistic Content for the Google Search Engine in Switzerland. *FehrAdvice & Partners AG*.

Kazdan, J., Schaeffer, R., Dey, A., Gerstgrasser, M., Rafailov, R., Donoho, D. L., & Koyejo, S. (2024). *Collapse or Thrive? Perils and Promises of Synthetic Data in a Self-Generating World* (arXiv:2410.16713). arXiv. https://doi.org/10.48550/arXiv.2410.16713

Khan, F. (2024, November 1). How OpenAI and Anthropic Are Cashing In on AI: A Look at Their Revenue Models. *Medium*. https://medium.com/@furqankhaan/how-openai-and-anthropic-are-cashing-in-on-ai-a-look-at-their-revenue-models-d9d9ae79dd28

Klemp, M., Rösch, K., Wagner, R., Quehl, J., & Lauer, M. (2023). LDFA: Latent diffusion face anonymization for self-driving applications. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3199–3205. https://openaccess.thecvf.com/content/CVPR2023W/E2EAD/html/Klemp_LDFA_Latent_Diffusion_Face_Anonymization_for_Self-Driving_Applications_CVPRW_2023_paper.html

Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. *International Conference on Machine Learning*, 1885–1894. http://proceedings.mlr.press/v70/koh17a?ref=https://githubhelp.com

La SPRE collecte la rémunération équitable pour les artistes-interprètes et les producteurs de phonogrammes. (n.d.). SPRE. Retrieved 12 March 2025, from https://www.spre.fr/

Laird, J. (2025, January 28). China's DeepSeek chatbot reportedly gets much more done with fewer GPUs but Nvidia still thinks it's 'excellent' news. *PC Gamer*. https://www.pcgamer.com/hardware/graphics-cards/chinas-deepseek-chatbot-reportedly-gets-much-more-done-with-fewer-gpus-but-nvidia-still-thinks-its-excellent-news/

L'ATLF a interrogé ses adhérents sur la post-édition. (2022). ATLF. https://atlf.org/latlf-a-interroge-ses-adherents-sur-la-post-edition/

Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés—Légifrance. (n.d.). Retrieved 25 November 2024, from https://www.legifrance.gouv.fr/loda/id/LEGISCTA000006095896

Lorphelin, V. (2024). *Un modèle économique pour valoriser le capital culutrel et humain par l'IA générative*. Controv3rse.

Loud and Clear by Spotify. (2023). Loud and Clear. https://loudandclear.byspotify.com/

Lutes. (2025). *Identifying the Economic Implications of Artificial Intelligence for Copyright Policy*.

Luzi, L., Mayer, P. M., Casco-Rodriguez, J., Siahkoohi, A., & Baraniuk, R. G. (2024). *Boomerang: Local sampling on image manifolds using diffusion models* (arXiv:2210.12100). arXiv. https://doi.org/10.48550/arXiv.2210.12100

Maleki, S., Tran-Thanh, L., Hines, G., Rahwan, T., & Rogers, A. (2014). *Bounding the Estimation Error of Sampling-based Shapley Value Approximation* (arXiv:1306.4265). arXiv. https://doi.org/10.48550/arXiv.1306.4265

Martínez, G., Watson, L., Reviriego, P., Hernández, J. A., Juarez, M., & Sarkar, R. (2023). *Combining Generative Artificial Intelligence (AI) and the Internet: Heading towards Evolution or Degradation?* (arXiv:2303.01255). arXiv. https://doi.org/10.48550/arXiv.2303.01255

Nasiripour, S., & Natarajan, S. (2019). Apple co-founder says Goldman's Apple card algorithm discriminates. *Bloomberg. Com*.

Nellis, S., & read, K. F. U. 2 min. (2025, January 27). *Nvidia says DeepSeek advances prove need for more of its chips*. Yahoo Finance. https://uk.finance.yahoo.com/news/nvidia-says-deepseek-advances-prove-192658104.html

« Non, l'intelligence artificielle ne remplacera pas les traducteurs et les traductrices! ». (2024, September 9). https://www.lemonde.fr/idees/article/2024/09/09/non-l-intelligence-artificielle-ne-remplacera-pas-les-traducteurs-et-les-traductrices_6308656_3232.html

Online Nation 2023 Report. (2023). Ofcom.

Peukert, C., Abeillon, F., Haese, J., Kaiser, F., & Staub, A. (2024). *Strategic Behavior and AI Training Data* (arXiv:2404.18445). arXiv. http://arxiv.org/abs/2404.18445

Pinaya, W. H. L., Tudosiu, P.-D., Dafflon, J., Da Costa, P. F., Fernandez, V., Nachev, P., Ourselin, S., & Cardoso, M. J. (2022). Brain Imaging Generation with Latent Diffusion Models. In A. Mukhopadhyay, I. Oksuz, S. Engelhardt, D. Zhu, & Y. Yuan (Eds.), *Deep Generative Models*

(Vol. 13609, pp. 117–126). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-18576-2_12

Racine, B. (2020). *L'auteur et l'acte de création*. Ministère de la culture.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695. http://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html

Sacra. (2024). *Anthropic revenue, valuation & growth rate*. https://sacra.com/c/anthropic/

Sacra. (2024). *OpenAI revenue, valuation & growth rate*. Open AI. https://sacra.com/c/openai/

Sadeghi, M., Arvanitis, L., Padovese, V., Pozzi, G., Badilini, S., Varcellone, C., Wang, M., Brewster, J., Huet, N., Fishman, Z., Pfaller, L., Adams, N., & Wollen, M. (2024). *Tracking AI-enabled Misinformation: Over 1100 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools*. NewsGuard. https://www.newsguardtech.com/special-reports/ai-tracking-center

Senftleben, M. (2024). *AI Act and Author Remuneration—A Model for Other Regions?* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4740268

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Skjuve, M. (2023). Why people use chatgpt. *Available at SSRN*, *4376834*.

SoA survey reveals a third of translators and quarter of illustrators losing work to AI - The Society of Authors. (2024, April 11). https://societyofauthors.org/2024/04/11/soa-survey-reveals-a-third-of-translators-and-quarter-of-illustrators-losing-work-to-ai/

Spotify launches revenue-sharing Partner Program. (2025, January 3). https://podnews.net/update/spotify-partner-program-launch

Thomas, A. (2025, February 24). *The AI licensing economy – CREATe*. https://www.create.ac.uk/blog/2025/02/24/the-ai-licensing-economy/

Thomas, P.-L. (2023, May 25). *Menacés par l'intelligence artificielle, les comédiens de doublage s'unissent dans un collectif international*. https://www.lemonde.fr/economie/article/2023/05/25/menaces-par-l-intelligence-artificielle-les-comediens-de-doublage-s-unissent-dans-un-collectif-international_6174777_3234.html

Trésor, D. générale du. (2024, December 5). *La chaîne de valeur de l'intelligence artificielle: Enjeux économiques et place de la France*. Direction générale du Trésor. https://www.tresor.economie.gouv.fr/Articles/2024/12/05/la-chaine-de-valeur-de-l-intelligence-artificielle-enjeux-economiques-et-place-de-la-france

Troyanskaya, O., Trajanoski, Z., Carpenter, A., Thrun, S., Razavian, N., & Oliver, N. (2020). Artificial intelligence and cancer. *Nature Cancer*, *1*(2), 149–152.

Update on KDP Title Creation Limits. (2023). https://www.kdpcommunity.com/s/article/Update-on-KDP-Title-Creation-Limits?language=en_US&forum=KDP%20Forum

Veselovsky, V., Ribeiro, M. H., & West, R. (2023). *Artificial Artificial Artificial Intelligence: Crowd Workers Widely Use Large Language Models for Text Production Tasks* (arXiv:2306.07899). arXiv. https://doi.org/10.48550/arXiv.2306.07899

Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports*, *13*(1), 15737.

Villalobos, P. (2022, November 10). *Will We Run Out of ML Data? Evidence From Projecting Dataset Size Trends*. Epoch AI. https://epochai.org/blog/will-we-run-out-of-ml-data-evidence-from-projecting-dataset

Vlasceanu, M., & Amodio, D. M. (2022). Propagation of societal gender inequality by internet search algorithms. *Proceedings of the National Academy of Sciences*, *119*(29), e2204529119.

Vulser, N. (2024, February 2). *Les traducteurs littéraires victimes de l'intelligence artificielle*. https://www.lemonde.fr/economie/article/2024/02/02/les-traducteurs-litteraires-victimes-de-l-intelligence-artificielle_6214361_3234.html

Wang, J. T., Deng, Z., Chiba-Okabe, H., Barak, B., & Su, W. J. (2024). *An Economic Solution to Copyright Challenges of Generative AI* (arXiv:2404.13964). arXiv. http://arxiv.org/abs/2404.13964

Wang, J. T., Mittal, P., Song, D., & Jia, R. (2024). *Data Shapley in One Training Run* (arXiv:2406.11011). arXiv. https://doi.org/10.48550/arXiv.2406.11011

Wang, S.-Y., Efros, A. A., Zhu, J.-Y., & Zhang, R. (2023). Evaluating data attribution for text-to-image models. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 7192–7203. http://openaccess.thecvf.com/content/ICCV2023/html/Wang_Evaluating_Data_Attribution_for_Text-to-Image_Models_ICCV_2023_paper.html

Wang, S.-Y., Hertzmann, A., Efros, A. A., Zhu, J.-Y., & Zhang, R. (2024). *Data Attribution for Text-to-Image Models by Unlearning Synthesized Images* (arXiv:2406.09408). arXiv. https://doi.org/10.48550/arXiv.2406.09408

Wenger, E. (2024). AI produces gibberish when trained on too much AI-generated data. *Nature*, *631*(8022), 742–743. https://doi.org/10.1038/d41586-024-02355-z