

INTELLIGENCE ARTIFICIELLE

« Tous les acteurs ont intérêt à la création d'un marché éthique et compétitif, respectueux de la chaîne de valeur. Un modèle européen »

La ministre de la Culture Rachida Dati a chargé Alexandra Bensamoun d'une mission sur la mise en œuvre du Règlement européen sur l'intelligence artificielle (IA), adopté en juin dernier. Le rapport, remis en décembre 2024, a été porté au niveau européen, dans le cadre de la réflexion menée par le Bureau de l'IA, en vue de l'élaboration d'un modèle de transparence sur les sources ayant servi à l'entraînement d'un modèle d'IA.

Semaine juridique, Édition générale : Quelles sont les raisons qui ont présidé à la mise en place de cette mission ?

Alexandra Bensamoun : La ministre de la Culture, Rachida Dati, est, depuis son arrivée, très engagée sur le sujet de l'intelligence artificielle. Il faut dire que le monde de la culture s'est unanimement et très fortement mobilisé, invoquant des raisons existentielles qui fragilisent l'ensemble des secteurs. En substance, le développement d'un modèle d'IA nécessite des données massives, qui servent à l'apprentissage. Parmi ces données, se trouvent des contenus protégés par le droit d'auteur et les droits voisins du droit d'auteur, utilisés sans autorisation ni rémunération.

Le rapport s'inscrit aussi dans le prolongement des réflexions, menées au sein de la Commission interministérielle de l'IA, qui a remis son rapport, « IA : notre ambition pour la France », au président de la République en mars 2024 (<https://www.elysee.fr/emmanuel-macron/2024/03/13/25-recommandations>

NDA : <https://www.culture.gouv.fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-propriété-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA/ia-et-transparence-des-données-d-entraînement-publication-du-rapport-d-alexandra-bensamoun-sur-la-mise-en-œuvre-du-règlement-européen-etabli-ant-Alexandra-Bensamoun-souhaite-ici-remercier-le-rapporteur,-Lionel-Ferreira,-pour-son-travail>



Entretien avec **ALEXANDRA BENSAMOUN**, professeure de droit, personnalité qualifiée au CSPLA (ministère de la Culture)

pour lia-en-France). Conformément au droit applicable, ce premier rapport réaffirme que l'utilisation de contenus protégés pour l'entraînement des modèles d'IA ne peut se faire que « dans le respect des droits de propriété intellectuelle » (V. *Recommandations clés*).

JCP G : L'article 53 du RIA dont la mission précise la portée vise les obligations incombant aux fournisseurs de modèles d'IA à usage général. Quelles sont-elles précisément dans le secteur culturel ?

A. B. : L'article 53 expose les obligations incombant aux fournisseurs de modèles d'IA

à usage général. S'agissant du domaine culturel, l'exigence est double. Le § 1, c), prescrit aux acteurs concernés de mettre en place une politique visant à se conformer à la législation de l'Union européenne en matière de droit d'auteur et de droits voisins. Le § 1, d), requiert l'élaboration et la mise à disposition du public d'un « résumé suffisamment détaillé du contenu utilisé pour entraîner le modèle (...) ».

Ces deux obligations s'imposent à tous les modèles, qu'ils soient ou pas publiés dans le cadre d'une licence libre et ouverte, et qu'ils présentent ou pas un risque systémique (lequel est présumé lorsque la puissance de calcul utilisée pour l'entraînement est supérieure à 10²⁵ FLOPS).

JCP G : Comment articuler les deux obligations en lien avec la culture ?

A. B. : La mission a considéré que les deux obligations participaient d'un même objectif : améliorer la transparence. En effet, le RIA crée pour les fournisseurs d'IA une obligation nouvelle et autonome de conformité à la racine (« *compliance by design* »), laquelle n'est effective que si la question de la politique de conformité et celle du résumé suffisamment détaillé sont traitées conjointement. Cette position est d'ailleurs soutenue par l'analyse contextuelle des dispositions. D'abord celles-ci sont unies par leur portée matérielle : elles sont

les seules au sein du § 1 à s'imposer à l'ensemble des fournisseurs d'IA, y compris ceux produisant des modèles libres (*cons. 104*). En outre, elles sont en partie liées dans leur esprit et dans leur régime, et d'ailleurs mentionnées comme une obligation unique (*cons. 108*). Aussi, la pertinence des informations du résumé suffisamment détaillé s'apprécie nécessairement à la lumière des mesures mises en œuvre par le fournisseur pour se conformer à ses obligations en droit d'auteur. La politique de conformité est en quelque sorte le négatif du résumé détaillé : ce que le second dit en plein, la première le dit nécessairement en creux.

JCP G : Que préconise la mission en ce qui concerne le résumé du contenu utilisé pour entraîner le modèle d'IA à usage général (53, § 1, d) du RIA) ? Et comment concilier l'exigence avec le secret des affaires ?

A. B. : La mission a considéré que le résumé ne saurait se borner à lister les principales sources de données, comme l'affirment certains acteurs de l'IA. En soutien, ceux-ci font valoir, d'une part, que le considérant 107 précise que pourraient figurer dans le résumé les « *principaux jeux ou collections de données utilisés pour entraîner le modèle, tels que les archives de données ou bases de données publiques ou privées de grande ampleur* » et, d'autre part, qu'il y a lieu de tenir compte « *de la nécessité de protéger les secrets d'affaires et les informations commerciales confidentielles* ».

Ce niveau d'information ne peut cependant permettre de donner au dispositif légal adopté un « effet utile », ainsi que le requiert souvent la CJUE. Il est en effet insuffisant pour atteindre l'objectif fixé par le législateur, à savoir « *aider les parties ayant des intérêts légitimes, y compris les titulaires de droits d'auteur, à exercer et à faire respecter les droits que leur confère la législation de l'Union* », comme y invite le considérant 107, qui mentionne par ailleurs les principaux jeux ou collections de données à titre d'exemples, et non de manière limitative.

Par ailleurs, si cette révélation ne doit évidemment pas compromettre le secret des affaires, l'invocation de celui-ci a des limites. En droit interne, l'article L. 151-7 du Code de commerce dispose que le secret des affaires ne peut être opposé aux autorités juridic-

tionnelles et administratives agissant, notamment, dans l'exercice de leurs pouvoirs d'enquête, de contrôle, d'autorisation ou de sanction. Et l'on notera avec intérêt que, dans l'affaire « *Dun & Bradstreet Austria GmbH* » (C-203/22) relative au traitement de données personnelles par une IA, ayant conduit à refuser la conclusion ou la prolongation d'un contrat de téléphonie mobile au motif que la personne ne présentait pas une solvabilité financière suffisante, l'avocat général a considéré, le 12 septembre 2024 (<https://curia.europa.eu/juris/document/document.jsf?text=&docid=290022&pageIndex=0&doclang=FR&mode=req&dir=&occ=first&part=1&cid=2434261>), que le secret des affaires ne pouvait conduire à écarter le droit qu'un individu tire du RGPD de comprendre comment une décision qui l'affecte est prise. Cette position paraît transposable aux droits qu'une personne tient des dispositions de droit d'auteur issues des textes européens. Le secret des affaires ne peut conduire, en vidant de toute substance le résumé suffisamment détaillé, à écarter le droit qu'un titulaire de droits tire du RIA à disposer d'éléments pouvant l'aider « *à exercer et à faire respecter les droits que leur confère la législation de l'Union* ». Enfin, la directive relative au secret des affaires envisage même l'hypothèse d'une règle de l'Union qui exigerait la révélation d'informations au public, y compris des secrets des affaires, pour des motifs d'intérêt public. Au-delà de ces arguments juridiques, le secret des affaires est encore délicat à opposer pour des entreprises qui utilisent pour l'entraînement le Common Crawl (répertoire mettant à disposition toutes les données du web) ou des bases de données partagées (comme BOOKS3 ou LAION).

Ainsi, une lecture finaliste et globale de l'article 53, § 1, d) donne toute sa portée à une expression qui semble être, de prime abord, un oxymore si l'on se concentre sur les seuls termes « résumé » et « détaillé ». Le résumé est donc suffisamment détaillé s'il permet d'atteindre la fin d'effectivité des droits. Autrement dit, le degré de détail s'apprécie au regard de l'objectif, avec une limite, celle du secret des affaires.

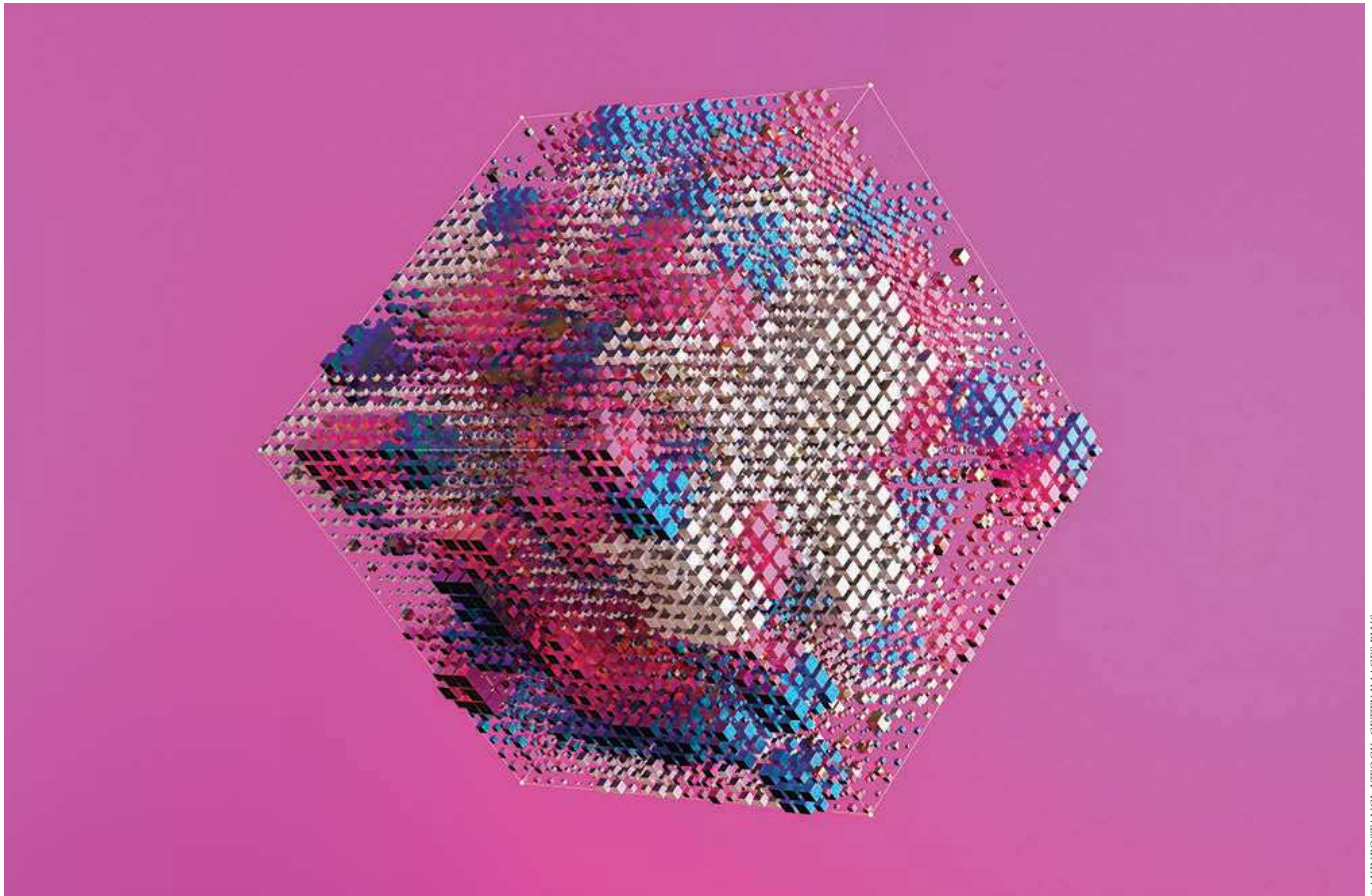
Dans ces conditions, le résumé doit être « *complet en termes de contenu* » (*cons. 107*), mais pas sur les techniques utilisées (tokenisation, filtrage...). La non-complétude évocatrice du terme « résumé » vise donc la recette et non les ingrédients !

JCP G : Cette obligation de transparence pourrait-elle constituer un frein à l'innovation dans le domaine de l'IA ?

A. B. : Cette opposition, très présente dans de nombreux discours, est plutôt étonnante pour un juriste. Elle suppose de présumer que le législateur européen poursuit avec le RIA une fin « anti-économique », alors même que sa démarche est fondée sur la construction du marché intérieur (*TFUE, art. 114*). Certes, le texte affiche également son ambition d'un niveau élevé de protection des droits fondamentaux. Mais pouvait-on accepter un autre positionnement ? L'approche par les risques adoptée, couplée à l'objectif d'harmoniser les conditions d'accès au marché de l'UE par renvoi à un mécanisme de conformité (conformément au Nouveau Cadre Législatif européen), permet un cadre de régulation équilibré. Sans compter que la vocation extraterritoriale du RIA soumet les acteurs hors UE aux mêmes règles dès lors qu'ils souhaitent pénétrer le marché européen.

La régulation permet encore de sécuriser le développement du marché, sous l'œil attentif des autorités de concurrence, conscientes des risques de dysfonctionnements et d'exclusions sur ce marché naissant (*V. par ex. ADLC, Avis 24-A-05 du 28 juin 2024, relatif au fonctionnement concurrentiel du secteur de l'intelligence artificielle générative*). L'incertitude juridique ne peut que conduire, comme c'est déjà le cas aux États-Unis (*Pour le droit d'auteur aux États-Unis, V. le recensement fait sur ce site : <https://chatgptiseatingtheworld.com/2024/08/27/master-list-of-lawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/>*), à la multiplication des contentieux ou à des transactions défavorables au développement du marché. Une régulation adaptée est sans doute nécessaire pour permettre l'innovation (*Anu Bradford, The False Choice Between Digital Regulation and Innovation, 6 oct. 2024, Columbia University*), pour que l'innovation soit synonyme de progrès.

Par ailleurs, la transparence est le gage de l'effectivité des droits. L'exercice des droits – de tous les droits – serait en effet gravement entravé par un recours systématique à l'opacité. Il est d'ailleurs intéressant de constater que les derniers textes récents sur le numérique – règlements européens sur les services numériques



©AKINROSTANLI/STOCK, GETTY IMAGES PLUS

(DSA) et sur les marchés numériques (DMA) – prônent la transparence.

Enfin, l'obligation de transparence devrait aussi conduire à traiter la problématique de la représentativité des données (biais d'entraînement pouvant, notamment, générer et amplifier des discriminations) et celle de la diversité des expressions culturelles ou du rayonnement de la culture française et francophone (V. Sommet de la Francophonie (oct. 2024) : Déclaration de Villers-Cotterêts, art. 20).

JCP G : Pensez-vous que la mise en œuvre de cette obligation de transparence puisse réellement favoriser la création d'un marché équitable pour la rémunération des contenus sous droits ?

A. B. : Il faut l'espérer. Les besoins des acteurs sont, en réalité, concordants. D'un côté, le développement d'un modèle d'IA nécessite des données. Mais les données du web ont déjà été utilisées (d'où la stagnation des versions récentes des modèles) et l'utilisation de données synthé-

« Le sujet de la rémunération est difficile mais essentiel pour apaiser les tensions. »

tiques pour l'entraînement conduit à la dégénérescence du modèle (Ilia Shumailov, Zakhar Shmaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson et Yarin Gal, *AI models collapse when trained on recursively generated data* : Nature, 24 juill. 2024). Les données de qualité – comme le sont les données culturelles et les contenus de presse – améliorent les performances et limitent les risques d'hallucinations. De l'autre côté, les titulaires de droits réclament le respect de leur monopole. Le risque est selon eux existentiels : alors même qu'ils sont exclus de la chaîne de valeur, les contenus générés sont susceptibles de concurrencer, voire de se substituer aux contenus humains.

En définitive, tous ont intérêt à la création d'un marché, un marché éthique et compétitif, respectueux de la chaîne de valeur. Un modèle européen.

Le sujet de la rémunération est difficile mais essentiel pour apaiser les tensions (V. la mis-

sion économique-juridique du CSPLA sur le sujet et la note d'étape juridique présentant les modèles de rémunération, déc. 2024 : <https://www.culture.gouv.fr/nous-connaître/organisation-du-ministère/Conseil-supérieur-de-la-proprière-littéraire-et-artistique-CSPLA/Travaux-et-publications-du-CSPLA/Missions-du-CSPLA/miission-relative-a-la-remuneration-des-contenus-culturels-utilises-par-les-systemes-d-intelligence-artificielle#:~:text=La%20mission%20relative%20%C3%A0%20la,Farchy%20pour%20le%20volet%20%C3%A9conomique>). Quelques contrats de licence ont déjà été passés. À l'heure du Sommet pour l'action sur l'IA (<https://www.elysee.fr/sommet-pour-l-action-sur-l-ia>), le soutien politique à l'ouverture d'un dialogue entre les acteurs serait un acte fort. ■

PROPOS RECUEILLIS PAR ÉLISE FILS