racine AVOCATS

Décryptage

Juillet 2025

Publication du modèle de résumé « *suffisamment détaillé* » des données d'entraînement des modèles d'IA à usage général

Le 24 juillet 2025, la Commission européenne a publié un avis explicatif avec, en annexe, un modèle de résumé destiné à guider les fournisseurs de modèles d'intelligence artificielle (« **IA** ») à usage général dans la rédaction du résumé du contenu utilisé pour l'entraînement de leurs modèles[1]. Ce modèle, découlant de l'article 53§1 (d) et du Considérant 107 du Règlement (UE) 2024/1689 du Parlement européen et du Conseil du 13 juin 2024 établissant des règles harmonisées concernant l'IA (« **AI Act** »), vise à garantir la transparence sur les sources d'entraînement des modèles d'IA à usage général.

Suivant cet article, les fournisseurs doivent élaborer et mettre « *à la disposition du public un résumé suffisamment détaillé du contenu utilisé pour entraîner le modèle d'IA à usage général, conformément à un modèle fourni par le Bureau de l'IA* ». La non-conformité avec cette obligation peut entraîner de lourdes amendes pouvant atteindre jusqu'à 15 000 000 d'euros ou 3 % du chiffre d'affaires annuel mondial total du fournisseur au cours de l'exercice précédent, le montant le plus élevé étant appliqué[2].

La publication de ce modèle intervient à point nommé, peu avant l'entrée en vigueur, le 2 août 2025, du chapitre V relatif aux « *Modèles d'IA à usage général* »[3]. Ce modèle complète le « *Code de bonnes pratiques de l'IA à usage général* » [4] et les *« Lignes directrices sur la portée des obligations incombant aux fournisseurs*

---

[1] Commission européenne, "*Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models*", Policy and legislation, 24 juillet 2025 : https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models
Commission européenne, "*Commission presents template for General-Purpose AI model providers to summarise the data used to train their model*", Press release, 24 juillet 2025 : https://digital-strategy.ec.europa.eu/en/news/commission-presents-template-general-purpose-ai-model-providers-summarise-data-used-train-their

[2] Article 101, AI Act

[3] Article 113, AI Act

[4] Commission européenne, « *Le code de bonnes pratiques de l'IA à usage général* », Politiques, 10 juillet 2025 : https://digital-strategy.ec.europa.eu/fr/policies/contents-code-gpai
L'on notera que ce code se compose de trois chapitres, dont deux consacrés à la transparence et au droit d'auteur. Ils offrent à l'ensemble des fournisseurs de modèles d'IA à usage général un moyen de démontrer leur conformité aux obligations prévues à l'article 53 de l'AI Act.

de modèles d'IA à usage général en vertu de la législation sur l'IA »[5], publiés respectivement les 10 et 18 juillet 2025.

La publication de ce modèle est l'occasion de revenir sur les apports de l'avis explicatif, ses finalités, le contenu attendu du résumé, les dérogations prévues, les obligations de mise à jour et les zones d'incertitude.

Finalité du résumé

L'avis explicatif rappelle la finalité du résumé, tel que précisé au Considérant 107 de l'AI Act :

- accroître la transparence concernant les données utilisées dans le cadre de l'entraînement préalable et de l'entraînement des modèles d'IA à usage général, y compris le texte et les données protégés par la législation sur le droit d'auteur ; et
- 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.

En particulier, s'agissant de cette deuxième finalité, les intérêts légitimes couvrent l'ensemble des droits protégés par le droit de l'Union qui devraient bénéficier d'une transparence renforcée. Il s'agit notamment des droits d'auteur et droits voisins, des droits des personnes concernées en matière de données, des droits fondamentaux à la non-discrimination, à la diversité linguistique et culturelle, ainsi que des libertés d'expression, d'information et de recherche scientifique[6].

À cette fin, comme le souligne l'avis explicatif à plusieurs reprises, les informations figurant dans le résumé doivent être « *généralement complet en termes de contenu plutôt que détaillé sur le plan technique* »[7] et présentées « *sous forme descriptive* »[8], simple et efficace[9].

Portée du « *contenu utilisé pour entraîner le modèle* » au sens de l'article 53§1(d) de l'AI Act

L'avis explicatif précise que les informations fournies dans le résumé doivent couvrir **l'ensemble des phases d'entraînement du modèle d'IA à usage général** - du pré-entraînement au post-entraînement - y compris les étapes d'alignement et de *fine-tuning*[10].

Le modèle du résumé publié fixe un socle commun minimal des informations à inclure dans le résumé, structuré en trois sections principales : « *informations générales* », « *liste des sources de données* » et « *aspects pertinents liés au traitement des données* »[11].

Il n'impose toutefois pas la divulgation des détails concernant les données et les œuvres précises utilisées pour l'entraînement du modèle, car cela irait au-delà de l'exigence prévue par l'article 53§1(d) de l'AI Act.

Protection des secrets d'affaires et des informations commerciales confidentielles

L'avis explicatif rappelle les préoccupations liées aux secrets d'affaires et aux informations commerciales confidentielles, telles que mentionnées au Considérant 107 de l'AI Act. Afin de trouver un équilibre entre

---

[5] Commission européenne, « *Lignes directrices sur la portée des obligations incombant aux fournisseurs de modèles d'IA à usage général en vertu de la législation sur l'IA* », Policy and Legislation, 18 juillet 2025 : https://digital-strategy.ec.europa.eu/fr/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act

[6] Avis explicatif, §6 à 12.

[7] Considérant 107, AI Act

[8] Considérant 107, AI Act

[9] Avis explicatif, §23.

[10] Avis explicatif, §13.

[11] Avis explicatif, §15.

transparence et protection des intérêts commerciaux, le document précise que les informations sensibles sur les sources de données et sur les modalités précises de sélection et d'entraînement des modèles peuvent rester confidentielles[12].

De manière non exhaustive, il prévoit que[13]:

- la divulgation des informations est limitée pour les données sous licence, dans la mesure où les titulaires de droits concernés sont déjà parties aux accords de licence[14] ;
- les ensembles de données privés, non couverts par une licence commerciale avec les titulaires de droits et obtenus auprès de tiers, ne doivent être listés que s'ils sont déjà connus du public (ou si le fournisseur souhaite les rendre publics)[15] ;
- pour les données synthétiques générées par des modèles d'IA utilisés à des fins d'entraînement et en particulier pour la distillation de modèles, les informations sont également limitées aux noms des modèles d'IA à usage général utilisés si ces modèles ont été mis sur le marché de l'Union ou, si d'autres modèles d'IA ont été utilisés, y compris les modèles appartenant aux fournisseurs, aux informations sur le modèle[16].

Modifications des modèles d'IA à usage général existants et mises à jour

Si un modèle d'IA à usage général déjà mis sur le marché de l'Union est modifié par un acteur en aval, les informations que ce dernier doit renseigner dans le résumé doivent se limiter aux contenus utilisés pour la modification du modèle. Le résumé doit également indiquer le nom du ou des modèles ayant fait l'objet de la modification[17].

En cas d'entraînement complémentaire du modèle, le résumé doit être mis à jour tous les six mois, ou plus tôt si les nouvelles données utilisées entraînent une modification substantielle du contenu du résumé, la première de ces échéances étant retenue[18].

Publication du résumé

Le résumé doit être rendu public, au plus tard, **au moment de la mise sur le marché du modèle d'IA à usage général dans l'Union**. Il doit être accessible sur le site Internet officiel du fournisseur de manière **claire et visible** ainsi que via tous les canaux de diffusion publique du modèle (par exemple, les plateformes en ligne)[19].

Pour les modèles d'IA à usage général déjà mis sur le marché **avant le 2 août 2025**, les fournisseurs devront prendre les mesures nécessaires pour publier le résumé correspondant, au plus tard, **le 2 août 2027**[20].

Prise en compte des PME

Une attention particulière est accordée aux petites et moyennes entreprises (PME), y compris les start-ups : celles-ci ne sont tenues de divulguer que les **5 % des noms de domaine** les plus explorés, ou **1 000 noms de**

[12] Avis explicatif, §18.
[13] Pour la liste complète, voir les §17 à 22 de l'avis consultatif.
[14] Avis explicatif, §19.
[15] Avis explicatif, §19.
[16] Avis explicatif, §21.
[17] Avis explicatif, §28.
[18] Avis explicatif, §29.
[19] Avis explicatif, §32.
[20] Avis explicatif, §33.

domaine (selon le nombre le plus faible) en fonction du volume de contenu extrait. Les autres fournisseurs doivent, quant à eux, divulguer les **10 % principaux**[21].

Zones d'ombre

Le modèle de résumé comporte encore certaines exigences ambiguës, dont la mise en œuvre pourrait s'avérer difficile en pratique. Par exemple, la section 2.5 de l'annexe intitulée « *Données synthétiques* » demande des **informations** — notamment une description générale des données d'entraînement — **concernant les modèles d'IA internes**, le cas échéant, utilisés pour générer des données synthétiques ayant servi à l'entraînement. Toutefois, le modèle ne précise pas clairement ce que doit contenir cette description générale.

Mise en perspective avec le rapport « *Template* » du CSPLA

Pour apprécier la portée du modèle publié par la Commission européenne, il est pertinent de le confronter à celui proposé par la mission « *template* » du Conseil supérieur de la propriété littéraire et artistique (« **CSPLA** ») en décembre 2024[22]. Tandis que la Commission européenne privilégie une approche centrée sur des informations générales, le CSPLA plaide pour une **transparence renforcée**, impliquant notamment l'identification précise des contenus non libres de droits utilisés. Pour chaque jeu de données, le modèle du CSPLA invite entre autres à détailler la méthode de moissonnage, la méthodologie pour assurer la conformité avec le droit de l'Union, les URLs moissonnées et leur date, ainsi que la taille et le type de données (images, son, multimodal, etc.), afin de permettre aux titulaires de droits d'exercer concrètement leurs prérogatives.

Néanmoins, même si le résumé exigé par la Commission européenne peut apparaître succinct, il ne saurait être exclu qu'un examen de fond puisse intervenir ultérieurement. La mission du CSPLA le relevait déjà en indiquant que le résumé constitue « *une sorte de contrôle de la **'recevabilité'**, vérifiant que le respect de la formalité, et non d'un examen substantiel, sur le fond* ». Elle ajoutait toutefois que « *Dans cette phase initiale de contrôle du respect de la conformité, **une évaluation « œuvre par œuvre » est exclue**, comme l'indique le considérant 108. Mais **il est possible qu'un examen sur le fond puisse intervenir dans un second temps**, notamment en cas de réclamation* »[23].

***

En publiant ce modèle de résumé, la Commission européenne poursuit son objectif de transparence encadrée pour les modèles d'IA à usage général, tout en veillant à préserver les intérêts légitimes des parties prenantes, notamment en matière de droits protégés et d'informations commerciales sensibles.

Bien que ce cadre apporte une première structuration bienvenue, certaines exigences demeurent imprécises et pourraient poser des difficultés pratiques lors de sa mise en œuvre. Des clarifications

---

[21] Avis consultatif, Annexe intitulée « *Modèle de résumé public de la formation pour les modèles d'IA à usage général requis par l'article 53§1(d) de l'AI Act* », voir Section 2.3. « *Data crawled and scraped from online sources* », p. 12.

[22] Conseil supérieure de la propriété littéraire et artistique (CSPLA), « *IA et Transparence des données d'entrainement : publication du rapport d'Alexandra Bensamoun sur la mise en œuvre du règlement européen établissant des règles harmonisées sur l'intelligence artificielle (RIA)* », Missions du CSPLA, 11 décembre 2024 : https://www.culture.gouv.fr/fr/nous-connaitre/organisation-du-ministere/Conseil-superieur-de-la-propriete-litteraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/ia-et-transparence-des-donnees-d-entrainement-publication-du-rapport-d-alexandra-bensamoun-sur-la-mise-en-aeuvre-du-reglement-europeen-etablissant

Conseil supérieure de la propriété littéraire et artistique (CSPLA), Rapport de mission relative à la mise en œuvre du règlement européen établissant des règles harmonisées sur l'intelligence artificielle (« *template* »), 11 décembre 2024 : https://www.culture.gouv.fr/fr/Media/medias-creation-rapide/cspla_rapport_ia_template_dec_.2024.pdf

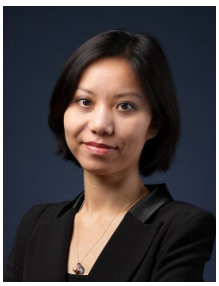
[23] CSPLA, Rapport Mission « *Template* », p. 24.

complémentaires seront sans doute nécessaires pour assurer une application harmonisée et efficace de ces nouvelles obligations, en particulier pour les PME et les acteurs en aval.

Auteurs

Charles Bouffier
Avocat associé
cbouffier@racine.eu
(+33) 6 64 10 87 36

Naomi Meynlé-Hamza
Juriste
nmeynle-hamza@racine.eu



Yibing Zhou
Avocate chinoise
yzhou@racine.eu