

Citer ses sources à l'ère des IAG

Les IA génératives (IAG) évoluant rapidement, les informations présentées ici sont susceptibles d'évoluer. N'hésitez pas à faire vous-même des tests pour confirmer ou infirmer les observations décrites.

Comment citer une IA ?

D'après Alexei Grinbaum, Directeur de recherche et président du Comité opérationnel pilote d'éthique du numérique du CEA, il faut « déclarer et expliciter l'usage qui [...] est fait [des outils IA] ». Si l'IA est « créatrice de solutions inédites », les découvertes « revien[nen]t [...] à celui qui a formulé le bon prompt -ou requête- et à ceux qui ont fabriqué le logiciel, un collectif d'auteurs regroupant les personnes ayant élaboré le corpus d'entraînement, fabriqué le modèle, etc... »¹.

En France, on peut **considérer de deux manières une œuvre créée par une IA** :

- si elle est **créée** seulement **avec l'assistance d'un ordinateur**, c'est la **personne humaine** à l'origine des actions pour la création qui **est titulaire des droits sur l'œuvre** ;
- si elle est **générée automatiquement**, sans intervention humaine, elle **ne pourra pas être qualifiée d'« œuvre »** car seule une personne humaine peut être considérée comme auteur et non une machine. A l'heure actuelle (février 2023), les créations issues d'un prompt ne sont pas protégées par le droit d'auteur français².

¹ Le Meur, H. et Grinbaum, A. (2024). IA : les filigranes au service de l'intégrité scientifique. Entretien avec Alexei Grinbaum. *Infolettre OFIS*, (10). <https://www.ofis-france.fr/wp-content/uploads/2024/02/EntretienOfis-AlexeiGrinbaum2024.pdf>

² Monneraye, L. de la. (2023, 15 février). L'intelligence artificielle peut-elle être auteure d'une œuvre ? Vaughan Avocats. <https://www.vaughan-avocats.fr/categories/revue-de-presse-12918/articles/lintelligence-artificielle-peut-elle-etre-auteure-dune-uvre-95188.htm>

Utiliser une IA pour rédiger ses références bibliographiques

Il est possible d'utiliser les IA génératives de texte pour rédiger ses références bibliographiques. Avant toute chose, il faut leur indiquer les règles de la norme et comment se construit une référence bibliographique. Un exemple peut leur être fourni. Cependant, une **relecture reste nécessaire**, car même avec toutes les informations nécessaires, la référence ne sera pas forcément construite correctement, et ce même quand on utilise des outils mettant en avant leur capacité à générer une référence normalisée comme <https://typeset.io/citation-generator>.

Lors de nos tests*, **ChatGPT-3.5, Copilot, Gemini et Perplexity** ont tous fourni des réponses erronées lorsqu'on leur demandait, sans leur donner d'exemple, de mettre en forme une référence bibliographique selon une norme donnée. Seul Le Chat de Mistral a fourni la réponse attendue.

Même avec un prompt détaillé indiquant le modèle de construction de la norme et les informations à faire figurer, toutes les réponses fournies ne sont pas exemptes d'erreurs. Un test a été fait avec la norme Vancouver :

- ChatGPT-3.5 et Gemini omettent d'abrégier le titre de la revue.
- Copilot inclut des mises en forme (gras, italique) inutiles et n'abrège pas non plus le titre de la revue.

Cette fois-ci, en plus du Chat de Mistral, Perplexity a également réussi l'exercice.

Enfin, aucun modèle testé n'a été capable de repérer l'ensemble des erreurs contenus dans une liste bibliographique et ce malgré le lien donné vers la norme.

*tests réalisés en mars et en avril 2024

Si vous disposez de ChatGPT-Plus (payant), il est possible de créer un chatbot dédié en lui fournissant la documentation sur un sujet (par exemple, la norme bibliographique) et en créant un prompt dédié. Cela permettra d'automatiser les requêtes futures.

En conclusion, à l'heure actuelle, **les IAG testées ne sont pas suffisamment performantes pour aider à la rédaction bibliographique, sauf si l'ensemble des éléments nécessaires à la rédaction leur sont détaillés et expliqués**. Zotero reste donc la meilleure alternative gratuite pour générer vos références selon une norme, même s'il ne dispense pas d'une relecture finale.

IA et respect du droit d'auteur

Les corpus servant à entraîner les IA génératives sont constitués de milliards de données (plus de **500 milliards pour GPT-3** par exemple³) issues du web. Dans le même ordre d'idée, le **générateur d'images Midjourney pourrait reproduire le style de 20 000 artistes**⁴.

Les sociétés développant les IA communiquent peu sur la source des données ayant servi à l'entraînement de leurs modèles. Ce manque de transparence conduit, outre-Atlantique, à la multiplication de procès initiés par des artistes graphiques (l'auteur et illustratrice de BD, Sarah Andersen), des écrivains (G.R.R. Martin, l'auteur de *Game of Thrones*) et même des journaux (le *New York Times*) pour non-respect du droit d'auteur⁵.

Face aux accusations, **OpenAI** (créateur de ChatGPT) **reconnait d'ailleurs ne pas pouvoir se passer de contenus sous droit d'auteur pour entraîner ses modèles**⁶. En **mars 2023**, un **accord pluriannuel entre OpenAI et le journal Le Monde** a été conclu, autorisant l'utilisation des contenus du journal pour l'entraînement de leur IAG ChatGPT. Mais ce type d'accord reste encore rare⁷.

Ainsi, lorsqu'on utilise un outil comme ChatGPT, il faut être conscient de ce problème et de son **incapacité à pouvoir sourcer ses réponses** et donc à potentiellement plagier.

D'autres IAG, comme **Perplexity ou Copilot**, peuvent **indiquer les sources** sur lesquelles elles appuient leurs réponses, mais celles-ci ne sont **pas toujours de qualité ou pertinentes**, surtout pour un travail universitaire.

³ Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901. <https://arxiv.org/pdf/2005.14165.pdf>

⁴ Oury, A. (2024, 13 février). Image ou texte, l'IA face à des procès en série. *Actualité*. <https://actualitte.com/article/115692/droit-justice/image-ou-texte-l-ia-face-a-des-proces-en-serie>

⁵ *Ibidem*

⁶ Bouhadjera, H. (2024, 9 janvier). ChatGPT : OpenAI admet ne pouvoir se passer de contenu soumis au droit d'auteur. *Actualité*. <https://actualitte.com/article/115130/legislation/chatgpt-openai-admet-ne-pouvoir-se-passer-de-contenu-soumis-au-droit-d-auteur>

⁷ https://www.lemonde.fr/le-monde-et-vous/article/2024/03/13/intelligence-artificielle-un-accord-de-partenariat-entre-le-monde-et-openai_6221836_6065879.html#

IA et fiabilité des sources

L'utilisation des IAG pour la recherche d'informations et de sources n'est pas conseillée et ce, pour diverses raisons :

- Le **modèle de fonctionnement** des IAG de texte (comme ChatGPT) est **basé sur la probabilité**. En tant que modèle de langage, il proposera ainsi le texte le plus probable en fonction des statistiques reliant les mots entre eux. Ainsi, **ChatGPT est un outil conversationnel, donc le but est de créer du texte, et non pas de fournir de l'information**.
- Une IAG de texte comme ChatGPT-3.5 a été entraînée sur un corpus délimité avant d'être mise à disposition du grand public. **Les données qui ont servi à l'entraînement ne sont pas renouvelées ou mises à jour**, ce qui rend impossible toute recherche sur un sujet d'actualité. Pour information, ChatGPT-3.5 a été entraîné avec des données jusqu'en janvier 2022 et la version payante ChatGPT-4 ou ChatGPT-Turbo, jusqu'en mars 2023. Les nouveautés dans ChatGPT font qu'aujourd'hui, pour certains sujets d'actualité, l'IAG précise qu'elle n'a pas d'informations. Cependant, ce n'est pas systématique⁸.
- Une IAG de texte comme ChatGPT est un outil conversationnel. Elle n'a pas été conçue pour effectuer une recherche documentaire. D'autres IAG fonctionnant grâce à un corpus d'entraînement achevé disposent d'un accès au web, ce qui permet une utilisation semblable à un moteur de recherche classique avec la possibilité de pouvoir faire des requêtes et de bénéficier de réponses en langage naturel. **Perplexity, par exemple, contrairement à ChatGPT-3.5, peut indiquer la source de ses propos et fournir des références**. Cependant, attention, malgré le développement de l'open access, de nombreux articles et ouvrages demeurent protégés par des paywall nécessitant une authentification préalable, et les IA ne peuvent donc pas y accéder.

Sur le site web : <https://fr.vittascience.com/ia/text.php>, il est possible de décortiquer le fonctionnement des IAG de texte : pour chaque mot répondu par l'IA, il est indiqué le pourcentage de probabilité qui a abouti à ce choix. Plus le pourcentage d'aléatoire sera élevé, plus l'IAG pourra faire preuve de créativité et s'affranchir des probabilités. Attention, quand ce pourcentage est trop élevé, la réponse apportée devient incompréhensible.

⁸ Lors d'une requête effectuée le 26 février 2024, nous avons demandé à ChatGPT-3.5 quelle pays avait remporté la coupe du monde de rugby en 2023. L'IAG a répondu, mais sa réponse était partiellement fausse (c'est bien l'Afrique du Sud qui a gagné, mais c'était contre la Nouvelle-Zélande et non l'Angleterre).

Comment protéger son œuvre des IA ?

Si l'œuvre est publiée, cela paraît très difficile. **L'ayant-droit peut refuser que son œuvre soit utilisée pour l'entraînement des IA mais il doit explicitement l'indiquer⁹ et mettre lui-même un moyen technique pour s'en assurer.** Cette barrière technologique est cependant quasiment impossible à mettre en place au vu de tous les canaux de diffusion possibles. En outre, les sociétés d'IAG n'étant pas toujours transparentes sur le détail des contenus utilisés lors des entraînements, **il est quasi impossible de déterminer si son œuvre a été utilisée ou non.**

Par contre, plusieurs institutions conseillent de **faire attention aux données ou éléments partagés lors de l'utilisation d'une IAG.** Les **prompts et les fichiers téléchargés dans les IA deviennent leur propriété** et peuvent être réutilisés pour fournir des réponses à d'autres utilisateurs¹⁰.

⁹ D'après Camille Jalicot, maître de conférences en droit de la propriété intellectuelle, une mention creative commons interdisant une réutilisation à des fins commerciales n'est pas suffisante pour interdire l'intégration d'un travail dans un corpus d'entraînement d'une IAG (Formation Urfist de Paris "Intelligence artificielle et propriété intellectuelle", 22 mars 2024)

¹⁰ Cancer Research UK. (2023, septembre). Guidance for researchers on the use of generative AI. https://www.cancerresearchuk.org/sites/default/files/cancer_research_uk_guidance_for_researchers_on_the_use_of_generative_ai.pdf?_gl=1*x0lqy8*_gcl_au*MTAzNTEzMzcwMy4xNzA3MzkyODE2*_ga*NDU4Mzg1MjE4LjE3MDE2OTYwMjE.*_ga_58736Z2GNN*MTcwNzkwODQxNy42LjEuMTcwNzkwMDA4Ny42MC4wLjA.&_ga=2.57057642.1146857737.1707908417-458385218.1701696021