L'Ofis
OFFICE FRANÇAIS
DE L'INTÉGRITÉ SCIENTIFIQUE

FAIT LE POINT

Février 2024

Systèmes d'intelligence artificielle générative : quelques points de vigilance

Depuis fin 2022, les systèmes d'intelligence artificielle (IA) générative du type de ChatGPT rencontrent un succès croissant. Le secteur de la recherche contribue amplement à cet engouement puisque les chercheuses et chercheurs se sont rapidement emparés de ces outils. L'Ofis souhaite attirer l'attention sur certains aspects de ces pratiques suceptibles de contrevenir aux exigences de l'intégrité scientifique.

De quoi s'agit-il ?

Les systèmes d'IA générative sont des outils capables de créer des contenus (du texte, des images, des graphiques...) en réponse à une requête (*prompt*) émise par l'utilisateur. Les dernières avancées dans le domaine des *large language models* (LLMs) permettent aujourd'hui à ces outils de générer des textes de très bonne qualité, et ce dans différentes langues et différents langages informatiques. ChatGPT, le bot d'OpenAI, est un des générateurs de texte les plus connus. Il en existe de nombreux autres: BLOOM du projet *BigScience,* Mixtral de l'entreprise française Mistral AI, Bard de Google, Bing chat de Microsoft etc. Comme il existe des IA génératrices d'images de plus en plus performantes : DALL.E , Midjourney. GPT image generator etc.

De l'assistance pour la rédaction d'articles à l'identification de pairs ou de références bibliographiques, nombreuses sont les pratiques pour lesquelles chercheurs et chercheuses peuvent recourir à l'aide des sytèmes d'IA générative.

Exemples de pratiques de recherche pour lesquelles l'IA générative peut être utilisée

- Définition de la question de recherche ou de l'hypothèse de départ
- Développement et argumentation de l'idée originale
- Sélection ou création de la méthodologie de recherche
- Assistance à la réalisation d'expérience, traitement d'images
- Résumé d'articles, support à la réalisation d'états de l'art
- Ecriture et optimisation de codes informatiques
- Rédaction, traduction, formatage de textes et d'images
- Réalisation de bibliographie, sélection d'articles
- Evaluation des projets ou des manuscripts, identification de pairs

Principes fondamentaux

Contrôle humain et responsabilité. Il est important de rappeler que les chercheuses et chercheurs ayant recours à des systèmes d'IA générative sont responsables des contenus générés qu'ils reproduisent dans leurs articles, commentaires ou autres productions de recherche - entre autres de leur fiabilité et de leur adéquation avec la règlementation en vigueur. Toute utilisation de ces outils exige le contrôle du résultat final par la personne responsable.

Transparence. Depuis juin 2023, le code de conduite européen pour l'intégrité scientifique[1] recommande la transparence : cacher l'utilisation d'IA ou d'outils automatisés dans la création de contenu ou dans la rédaction de publications y est désormais considéré comme un manquement à l'intégrité scientifique.

Points de vigilance pour l'intégrité scientifique

L'Ofis attire aussi l'attention de ceux et celles qui utilisent ces outils sur quelques points de vigilance. Etant données la vitesse des avancées en IA et l'évolution en cours de la règlementation, cette liste est appelée à évoluer.

Fiabilité Les systèmes d'IA génératives font parfois des erreurs, et présentent de manière très vraisemblable des informations erronées voire complètement inventées (couramment appeler des "hallucinations").[2] Cela a par exemple été mis en évidence dans la génération d'états de l'art et de références bibliographiques,[3] ou de réponses à des questions scientifiques.[4] Cela expose les chercheuses et chercheurs à un risque de diffuser de fausses informations, voire à de la fabrication et de la falsification.

Propriété Il est possible que le contenu généré par les systèmes d'IA générative soit issu de données d'entrainement protégées par un *copyright*.[5] Il existe dont un risque de plagiat par l'utilisateur qui s'approprie ce contenu. Par ailleurs, un système d'IA générative ne peut pas être reconnu comme auteur d'un article ou d'autres productions de recherche (lire « ChatGPT, auteur de publication scientifique ? »). Les détails de l'utilisation des LLMs doivent être déclarés dans les méthodes et/ou les remerciements.

Confidentialité Les systèmes actuels d'IA générative n'offrent pas une protection suffisante en matière de données personnelles – qu'il s'agisse de la protection des données d'entrainement ou de celles qui proviennent des requêtes des utilisateurs.[2,5] Ces systèmes présentent un risque de violation de plusieurs dispositions du RGPD, telles que la confidentialité, le consentement ou le droit à l'oubli.[2,5] Les chercheuses et chercheurs doivent donc éviter de partager des données personnelles ou confidentielles dans leurs requêtes.

[1] ALLEA, *The European Code of Conduct for Research Integrity*, 2023
[2] CNPEN, Avis n°7, "Systèmes d'intelligence artificielle générative : enjeux d'éthique", 2023
[3] W. H. Walters et E. I. Wilder, « Fabrication and errors in the bibliographic citations generated by ChatGPT », *Sci Rep*, vol. 13, n° 1, Art. n° 1, sept. 2023, doi: 10.1038/s41598-023-41032-5.
[4] H. Zheng et H. Zhan, « ChatGPT in Scientific Writing: A Cautionary Tale », *The American Journal of Medicine*, mars 2023, doi: 10.1016/j.amjmed.2023.02.011.
[5] Novelli, Claudio *et al*.. « Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity ». 2024. arXiv. https://doi.org/10.48550/arXiv.2401.07348.