

De la méthode à l'intégrité : gouvernance éthique et légale d'un outil d'IA générative pour la recherche « RedactoSci »

Paulin Rodrigue Njayou Tchapda, Sarra Yasmine Bali, Rachidatou Mabey Insa et Fatimata Zahra Diop

Université du Québec en Outaouais

9 Décembre 2025, Automne 2025

Plan de la présentation

- 1 Introduction
- 2 Contexte et problématique
- 3 Objectifs
- 4 Méthodologie (Rappel)
- 5 Analyse : intégrité, éthique, ÉDI
 - Analyse éthique
 - Aspects ÉDI
- 6 Propriété intellectuelle
- 7 Valorisation et diffusion
- 8 Conclusion

Introduction et justification du sujet

- Expansion fulgurante de l'IA générative en 2025 : texte, images, code, audio.
- Les LLM permettent de produire en quelques secondes des contenus complexes [1].
- Les milieux universitaires sont transformés : rédaction, analyse, recherche documentaire.
- Notre équipe propose **RedactoSci**, un modèle spécialisé formé sur millions d'articles.

Mais : intégrité, plagiat, hallucinations, propriété intellectuelle et gouvernance deviennent critiques [2].

Impact de l'IA générative sur la recherche

- Automatisation croissante des tâches rédactionnelles.
- Forte dépendance possible aux outils d'IA.
- Risques : erreurs factuelles, dilution de l'analyse humaine [2].
- Besoin de mécanismes institutionnels pour garantir une utilisation responsable.

Problématique

Les chercheurs font face à :

- surcharge informationnelle [1].
- délais serrés,
- syndrome de la page blanche,
- maintien difficile de l'intégrité scientifique.

RedactoSci a été conçu pour répondre à ces défis.

Objectifs SMART de RedactoSci

- **S** : assister la rédaction scientifique.
- **M** : réduire de 30% le temps de production.
- **A** : automatiser tâches : plans, résumés, reformulation, code.
- **R** : faciliter l'accès aux ressources fiables.
- **T** : déploiement sur un semestre universitaire.

Rappel du schéma méthodologique

- Collecte de données scientifiques : articles, rapports, thèses.
- Nettoyage : suppression doublons, correction erreurs, normalisation.
- Modèle GPT-like basé sur Transformer [3].
- Fine-tuning : résumés, concepts, reformulations, suggestions bibliographiques.
- Évaluations intermédiaires + retours utilisateurs.

Analyse d'intégrité et d'éthique

Trois enjeux majeurs :

- ① **Plagiat algorithme** : reproduction involontaire de données d'entraînement [4].
- ② **Hallucinations** : production d'informations fausses [5].
- ③ **Perte de compétence** : dépendance excessive et diminution des capacités critiques [6].

- Les LLM peuvent mémoriser et regénérer du contenu protégé.
- **Solutions RedactoSci :**
 - reformulation automatique,
 - détection des formulations proches,
 - rappel des obligations de citation.

Fabrication de données (Hallucinations)

- Contenu plausible mais faux ou inventé.
- **Solutions :**
 - système de détection d'incertitude,
 - avertissements,
 - encouragement à vérification manuelle.

Perte de compétence (Deskilling)

- Risque de réduire la pensée critique et l'analyse autonome.
- **RedactoSci vise à :**
 - renforcer l'apprentissage actif,
 - fournir explications méthodologiques,
 - éviter remplacement du raisonnement humain.

Aspects ÉDI

Aspect	Description et rôle de <i>RedactoSci</i>
Biais linguistiques	La recherche se fait avec la langue anglais majoritairement. <i>RedactoSci</i> intègre des données multilingues pour favoriser la diversité scientifique.
Biais culturels et de genre	Le langage scientifique peut reproduire des stéréotypes. <i>RedactoSci</i> ajuste les formulations fournies pour assurer le respect et l'inclusion.
Transparence et inclusion dans la conception	Les données sont documentées et traçables. <i>RedactoSci</i> assure la fiabilité des données et la transparence.

Gestion de la propriété intellectuelle

- L'IA n'est pas considérée comme auteur légal.
- L'utilisateur humain demeure responsable du contenu.
- Obligation de divulguer l'usage de l'IA.
- Protection des données : éviter prompts sensibles, anonymisation.

- **Science ouverte** : transparence, DOI, code public, model card.
- **Secret commercial** : API, NDA, modules propriétaires.
- **Solution recommandée : modèle hybride** :
 - version gratuite = fonctions essentielles,
 - version payante = fine-tuning, analyses avancées.

Plan de valorisation

Trois options :

- création d'une spin-off universitaire,
- service gratuit pour maximiser l'impact scientifique,
- **modèle hybride** (recommandé).

Diffusion aux éditeurs scientifiques

- Respect normes IEEE et Springer.
- Transparence sur l'usage de l'IA.
- Résultats reproductibles et vérifiables.

Diffusion aux comités d'éthique

- Démarche conforme à protection des données.
- Fourniture des protocoles et documents requis.
- Mise en évidence des bénéfices pédagogiques.

Conclusion

- RedactoSci répond à des défis méthodologiques majeurs.
- Importance d'une gouvernance éthique rigoureuse.
- Le modèle hybride concilie accessibilité et durabilité.
- L'IA peut soutenir la recherche si elle est utilisée de manière responsable.

Références I

- [1] M. Humlund and C. Zhang, "Generative ai's impact on graduate student professional writing productivity and quality," *International Journal of Artificial Intelligence in Education*, 2025.
- [2] UQAM, "Intégrer l'ia dans la recherche universitaire."
<https://collimateur.uqam.ca/collimateur/integrer-lia-dans-la-recherche-universitaire/>, 2025.
- [3] G. Antonesi, T. Cioara, I. Anghel, V. Michalakopoulos, E. Sarmas, and L. Toderean, "From transformers to large language models : A systematic review of ai applications in the energy sector towards agentic digital twins," *arXiv preprint arXiv:2506.06359*, 2025.

Références II

- [4] N. Carlini, F. Tramer, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. Brown, D. Song, U. Erlingsson, *et al.*, "Extracting training data from large language models," in *30th USENIX Security Symposium (USENIX Security 21)*, pp. 2633–2650, 2021.
- [5] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, *et al.*, "A survey on hallucination in large language models : Principles, taxonomy, challenges, and open questions," *ACM Transactions on Information Systems*, vol. 43, no. 2, pp. 1–55, 2025.
- [6] M. Gerlich, "Ai tools in society : Impacts on cognitive offloading and the future of critical thinking," *Societies*, vol. 15, no. 1, p. 6, 2025.