INF5163 – Méthodologie de recherche en informatique

Examen Finale, Groupe 8

*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

25 Novembre, Session Automne 2025

Table des matières

Liste des tableaux

Table des figures

1 Introduction et justification du sujet

2 Problématique et objectifs

3 Rappel du schéma méthodologique

Le modèle RedactoSci a été développé en suivant une méthodologie rigoureuse, inspirée des pratiques courantes en intelligence artificielle appliquée à la recherche scientifique. Elle repose sur trois étapes clés : la sélection des données d'entraînement, le choix d'une architecture de modèle de langage moderne et l'adaptation du modèle au contexte académique.

3.1 Collecte et préparation des données

La diversité et la qualité des données constituent des éléments essentiels à la performance des modèles de langage de grande taille *(LLM)*[1]. En nous basant sur cette idée, la première étape de notre méthodologie consiste à définir un ensemble de textes scientifiques destiné à l'entraînement du modèle. Nous avons défini un ensemble composé de :

— Des articles scientifiques majoritairement récents.

— Des rapports techniques.

— Des livres scientifiques et chapitres d'ouvrages.

— Des thèses de mémoire universitaires.

Pour assurer la fiabilité des données, nous avons appliqué un processus de nettoyage incluant la suppression des doublons, la correction des erreurs typographiques, et la normalisation du formatage. on a également assurer la fiabilité des ressources utilisées en privilegiant des sources reconnues et en virifiant la provenance des documents. Les modèles entraînés sur des données mal filtrées présentent un risque accru d'erreurs factuelles et d'incohérences dans leurs réponses. Cette observation justifie l'importance d'un contrôle rigoureux de la qualité des données lors de la phase de préparation.

3.2 Choix de l'architecture du modèle

Les modèles GPT-like restent parmi les plus performants dans la production de textes spécialisés. Pour concevoir RedactoSci, nous avons utilisé un modèle pré-entraîné de type GPT-like, fondé sur l'architecture *Transformer*. L'architecture *Transformer* a fondamentalement révolutionné le traitement automatique du langage naturel et est devenue la pierre angulaire des modèles de langage de grande taille modernes [2]. Le choix du *Transformer* permet donc notre modèle de disposer d'une base solide et performante pour générer des textes cohérents et pertinents.

3.3 Méthode adaptation au contexte académique (Fine-tuning)

Afin d'adapter le modèle GPT-like aux exigences scientifiques. Nous avons applique une méthode d'instruction-based fine-turning. Cette approche consiste à fournir au modèle des exemples structurés tels que :

— Des résumés d'articles.

— Des explications des différents concepts scientifiques.

— Des introductions et des conclusions bien structurés.

— Des reformulations de textes complexes en langage plus accessible.

— Des suggestions de références bibliographiques pertinentes.

Grace à ce processus, RedactoSci apprend à reproduire la structure, la rigueur et le style attendus dans les publications scientifiques. Ce type de fine-tuning cible permet d'obtenir un LLM spécialisé, performant dans un domaine précis[3]

4 Analyse d'intégrité, éthique et ÉDI

4.1 Analyse d'intégrité et d'éthique

L'utilisation d'un modèle de langage dans un contexte académique soulève plusieurs enjeux liés à l'intégrité scientifique. Nous présentons ici trois risques majeurs : le plagiat algorithmique, la fabrication de données (hallucinations) et la perte de compétence des utilisateurs.

a) **Plagiat algorithmique**
L'un des risques les plus importants associés aux grands modèles de langage est leur capacité à reproduire mot pour mot des extraits issus de leur ensemble d'entraînement. [4]. Les LLM peuvent mémoriser et restituer des passages entiers lorsque certaines séquences rares ou reconnaissables leur sont présentées[4]. Ce phénomène crée un risque élevé de plagiat si les données d'entraînement contiennent des œuvres protégées par le droit d'auteur.

Pour réduire ce risque, *RedactoSci* intègre des mécanismes favorisant la reformulation et décourageant la génération de contenu trop proche de sources existantes. De plus, des vérifications internes détectent les formulations suspectes et avertissent l'utilisateur lorsque le texte généré est trop similaire à un contenu potentiellement protégé. L'utilisateur est également encouragé à citer explicitement les sources qu'il utilise réellement, afin de respecter les principes d'intégrité académique.

b) **Fabrication de données (hallucinations)**
Même les modèles de langage les plus avancés continuent d'inventer des faits, des références ou des entités fictives, c'est ce qu'on appelle l'hallucination. Les grands modèles de langage (LLM) génèrent parfois un contenu plausible mais les hallucinations demeurent un défi majeur à mesurer que ces modèles sont de plus en plus utilisés dans des contextes réels et à fort enjeu[5].

Pour atténuer ce problème, *RedactoSci* inclut des mécanismes de *détection d'incertitude* permettant au modèle de signaler les réponses dont la fiabilité est limitée. Lorsqu'une donnée ne peut être confirmée, le modèle propose des formulations prudentes ou invite explicitement l'utilisateur à effectuer des vérifications manuelles.

c) **Perte de compétence (deskilling)**
La dépendance à long terme à l'égard des outils d'IA pour l'externalisation cognitive pourrait également éroder des compétences cognitives essentielles telles que la mémoire, l'analyse et la résolution de problèmes. À mesure que les individus s'appuient davantage sur l'IA, leurs capacités cognitives internes risquent de s'atrophier, entraînant une diminution de la mémoire et de la santé cognitive à long terme[6].

RedactoSci a été conçu pour accompagner l'utilisateur plutôt que le remplacer. Il fournit des explications, des pistes de réflexion et des justifications méthodologiques plutôt que des réponses complètes et définitives. Cette approche favorise l'apprentissage actif et encourage l'utilisateur à développer ses propres compétences.

4.2 Aspects ÉDI

RedactoSci vise à promouvoir l'équité, la diversité et l'inclusion (ÉDI) selon plusieurs axes opérationnels et éthiques. Conçu comme un modèle de langage scientifiques, il s'appuie sur des données de différentes langues, cultures et discipline. Son but est de réduire les biais structurels dans les systèmes d'intelligence artificielle.

a) **Biais linguistiques**

La plupart des modèles de langage sont entrainés sur des travaux et textes en anglais, ce qui peut réduire la visibilité de la recherche en langue différentes. Cela représente un déséquilibrement qui peut créer une dépendance à la langue dominante et réduit la diversité des idées. Le modèle *RedactoSci* cherche à répondre à ces besoins et corriger ce déséquilibre en s'appuyant sur des données multilingues. Grace a cette approche, la recherche multilingue peut bénéficier de la même précision et que celle produite en anglais.

b) **Biais culturels et de genre**

Le choix des mots dans les sujets scientifiques influence fortement la crédibilité et réduit la valeur de certaines contributions. Pour limiter ce type d'erreurs, *RedactoSci* a été conçu pour ajuster les formulations. Le modèle aussi adopte une approche fondée sur le respect, l'équité et la sensibilité culturelle. Son objectif est de garantir que chaque perspective puisse être exprime de manière équilibrée, respectueuse et sans préjugé.

c) **Transparence et inclusion dans la conception**

L'un des principes fondamentaux de notre modèle est la transparence. Tous les documents et les informations utilisé pour l'apprentissage de *RedactoSci* sont bien choisit et documentés afin de préciser leur contexte d'utilisation et leur origine. De plus il permet d'identifier et corriger les déséquilibres dans les données dont d'assurer la qualité scientifique du modèle. Le modèle adopte aussi une approche inclusive, où chaque étape prend en considération la diversité linguistiques et culturelles Cela garantit que le modèle reste pertinent et fidèle aux valeurs de diversité que la recherche universitaire doit défendre.

5 Gestion et propriété intellectuelle

6 Valorisation et Diffusion

7 Contributions individuelles

Références

[1] H. Chen, A. Waheed, X. Li, Y. Wang, J. Wang, B. Raj, and M. I. Abdin, "On the diversity of synthetic data and its impact on training large language models," *arXiv preprint arXiv :2410.15226*, 2024.

[2] 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.

[3] A. Narayan, M. F. Chen, K. Bhatia, and C. Re, "Cookbook : A framework for improving llm generative abilities via programmatic data generating templates," *arXiv preprint arXiv :2410.05224*, 2024.

[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.