الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي -8 ماي 1945- سيدي بلعباس
École Supérieure en Informatique -08 Mai 1945- Sidi Bel Abbès

Thesis

To obtain the diploma of Master
Field : Computer Science
Specialty : Information system and web developement

Theme

Ai-Generated Text Detection & Sentiment analysis: State of the Art

---

Presented By : Yahiaoui Meriem

Submitted : September 28, 2025
In front of the jury composed of :

Dr. CHAIB Souleyman	President
Pr.BENSLIMANE Sidi Mohammed	Supervisor
Dr.BENABDERRAHMANE Sid Ahmed	Co-supervisor
Dr. CHEIKH Asma	Examinator

Academic year : 2024/2025

ABSTRACT

The rapid evolution of large language models has transformed the landscape of natural language generation, enabling the production of highly coherent and human-like text at unprecedented scale. While these advancements have brought remarkable benefits in productivity, accessibility, and automation, they also introduce profound societal risks. The ability of large language models to generate persuasive and context aware narratives facilitates the spread of misinformation, manipulation of public opinion, and intensification of political or social tensions particularly during conflicts and crises. Detecting such content poses significant challenges due to its fluency, contextual relevance, and capacity to mimic human expression, which often render it indistinguishable from authentic communication. Existing moderation tools struggle to address issues such as adversarial manipulation, domain specific narratives, and multilingual adaptation, making the development of robust detection mechanisms both urgent and complex. This dual nature of large language models capable of both constructive applications and malicious exploitation underscores the need for effective, adaptable, and transparent safeguards.

In this thesis we provide an overview of the state of the art in AI-generated text detection and sentiment analysis.It synthesizes recent research findings, identifies key advancements, and examines the current limitations and challenges in the field. The review highlights trends shaping future developments, including efforts to improve accuracy, scalability, and resilience, while ensuring adaptability across different contexts and languages.

Keywords **:** Natural Language Processing, Large Language Models,Deep Learning,Machine Learning, Sentiment Analysis,text classification, Misinformation, Social Media,Transformers

RÉSUMÉ

L'évolution rapide des grands modèles de langage a transformé le paysage de la génération automatique de texte, en permettant la production d'écrits hautement cohérents et proches du langage humain à une échelle sans précédent. Bien que ces avancées aient apporté des bénéfices remarquables en matière de productivité, d'accessibilité et d'automatisation, elles introduisent également des risques sociétaux profonds. La capacité des grands modèles de langage à générer des récits persuasifs et sensibles au contexte facilite la diffusion de la désinformation, la manipulation de l'opinion publique et l'intensification des tensions politiques ou sociales, en particulier lors de conflits et de crises. La détection de tels contenus pose des défis importants en raison de leur fluidité, de leur pertinence contextuelle et de leur aptitude à imiter l'expression humaine, ce qui les rend souvent indiscernables de communications authentiques. Les outils de modération existants peinent à traiter des problématiques telles que la manipulation adversariale, les récits spécifiques à un domaine ou encore l'adaptation multilingue, rendant le développement de mécanismes de détection robustes à la fois urgent et complexe. Cette double nature des grands modèles de langage capables à la fois d'applications constructives et d'usages malveillants souligne la nécessité de garde-fous efficaces, adaptatifs et transparents.

Dans ce mémoire, nous présentons un état de l'art de la détection de textes générés par intelligence artificielle et de l'analyse de sentiment. Il synthétise les travaux récents, identifie les avancées clés et examine les limites ainsi que les défis actuels du domaine. La revue met en évidence les tendances qui orientent les développements futurs, notamment les efforts visant à améliorer la précision, l'évolutivité et la robustesse, tout en garantissant une adaptabilité aux différents contextes et langues.

***Mots-clés* :** Traitement automatique du langage naturel, Grands modèles de langage, Apprentissage profond, Apprentissage automatique, Analyse de sentiment, Classification de texte, Désinformation, Réseaux sociaux, Transformeurs

ملخص

لقد غيّر التطور السريع للنماذج اللغوية الكبيرة مشهد توليد اللغة الطبيعية، حيث أتاح إنتاج نصوص شديدة الترابط وقريبة من اللغة البشرية على نطاق غير مسبوق. وبينما جلبت هذه التطورات فوائد ملحوظة في مجالات الإنتاجية وسهولة الوصول والأتمتة، فإنها أدخلت أيضًا مخاطر اجتماعية عميقة. إذ تتيح قدرة النماذج اللغوية الكبيرة على توليد سرديات مقنعة وواعية بالسياق انتشار المعلومات المضللة، والتلاعب بالرأي العام، وتصعيد التوترات السياسية أو الاجتماعية، لا سيما أثناء النزاعات والأزمات. وتشكل عملية كشف مثل هذا المحتوى تحديات كبيرة نظرًا لطلاوته وملاءمته السياقية وقدراته على تقليد التعبير البشري، مما يجعله في كثير من الأحيان غير قابل للتمييز عن التواصل الأصيل. وتعجز أدوات المراقبة الحالية عن معالجة قضايا مثل التلاعب العدائي، والسرديات الخاصة بالمجالات، والتكيّف متعدد اللغات، مما يجعل تطوير آليات كشف قوية أمرًا عاجلاً ومعقدًا في آن واحد. إن هذه الطبيعة المزدوجة للنماذج اللغوية الكبيرة ـ القادرة على كل من التطبيقات البنّاءة والاستخدامات الخبيثة تبرز الحاجة إلى ضمانات فعّالة وقابلة للتكيّف وشفافة.

في هذه الأطروحة، نقدم نظرة شاملة حول أحدث ما توصلت إليه الأبحاث في مجال كشف النصوص المولّدة بالذكاء الاصطناعي وتحليل المشاعر. فهي تلخص نتائج البحوث الحديثة، وتحدد أبرز التطورات، وتستعرض القيود والتحديات الحالية في هذا المجال. كما تسلط المراجعة الضوء على الاتجاهات التي تشكّل التطورات المستقبلية، بما في ذلك الجهود المبذولة لتحسين الدقة، وقابلية التوسع، والمرونة، مع ضمان القدرة على التكيف مع السياقات واللغات المختلفة.

الكلمات المفتاحية: معالجة اللغة الطبيعية، النماذج اللغوية الكبيرة، التعلم العميق، التعلم الآلي، تحليل المشاعر، تصنيف النصوص، المعلومات المضللة، شبكات التواصل الاجتماعي، المحوّلات