

IDIOMA: FRANCÊS

Área 4

* Indica uma pergunta obrigatória

1. E-mail *

2. ÁREA *

Marcar apenas uma oval.

☐

4 - Linguística, Letras e Artes

3. NOME DO CANDIDATO *

4. NÚMERO DA INSCRIÇÃO *

5. NÚMERO DO CPF *

Leia o texto e responda as questões a seguir em Português. Todas as questões devem ser respondidas de acordo com o texto. As respostas digitadas neste formulário eletrônico constituirão o ÚNICO documento válido para correção da prova.

Comment « dé-IA-iser » nos écrits pour éviter la disparition des particularités des langues ?

Publié le 7 juillet 2026

L'intelligence artificielle (IA) générative présente un certain nombre de [tics de langage], des mots et tournures qu'elle suremploie par rapport aux rédacteurs humains. Une étude estime qu'au moins 13,5 % des publications scientifiques de 2024 ont fait intervenir une IA dans leur rédaction, un chiffre qui grimpe à 40 % dans certaines revues. Claude, ChatGPT ou Deepseek sont nourries des mêmes sources (Web, livres, Wikipédia, code...). Ces grands modèles de langage (LLM) reposent sur la prédiction statistique du prochain token (mot ou fragment de mot) le plus probable dans une séquence donnée.

L'IA sélectionne donc systématiquement les formulations les plus accessibles et conventionnelles disponibles dans ses données d'entraînement. Elle écrit en perdant l'originalité des styles humains, leur irrégularité, leur diversité. Elle perd également le double sens, voire la faculté critique. Une expérience menée auprès de 293 participants (« *writers* ») invités à écrire une courte histoire a montré une augmentation de la similarité entre histoires conçues avec l'IA générative et donc une perte de créativité collective. Chacun écrit « mieux », mais tout le monde finit par écrire pareil.

Au-delà de la génération de tokens, plusieurs éléments peuvent expliquer les tics de langage de l'IA. La première cause tient à l'étape dite d'« alignement » : une fois entraîné sur des milliards de textes, le modèle est ajusté par des humains qui notent ses réponses, un procédé connu sous le nom de Reinforcement Learning from Human Feedback (RLHF). Des annotateurs humains évaluent le caractère neutre et inoffensif des réponses du modèle. Ils retiennent les réponses polies, structurées présentant des phrases relativement courtes et des mots de liaison donnant le sentiment d'une pensée logique.

La deuxième tient au terrain d'entraînement des IA. Alimentées de contenus marketing tirés du Web ou de documents institutionnels, les IA sont soit emphatiques, soit terriblement ennuyeuses et reproduisent les tics langagiers présents dans ces corpus. Les recherches montrent une prédominance des domaines commerciaux (.com) largement présents dans l'un des grands jeux de données de référence pour entraîner les modèles de langage, le Colossal Clean Crawled Corpus ou C4. Il existe, également, une forte prévalence de contenus dupliqués, synthétiques ou de basse qualité. C'est ce matériau standardisé issu du Web marchand anglo-saxon que le modèle, statistiquement, apprend à imiter. Mais que révèlent ces tics de langage ? Si les tics sont en train de coloniser le Web, ils menacent les IA elles-mêmes par des phénomènes de contamination croisée.

Vers la fin des langues régionales ?

À terme se profile la fin des langues régionales inconnues des IA, mais également la fin de l'apprentissage des langues. Au-delà des tics de langage, les IA imposent sur tous les réseaux sociaux des traductions automatisées dans la langue de l'utilisateur. Objectif : lui permettre de lire les contenus du monde entier sans barrières linguistiques, avec le risque de véhiculer une langue traduite pauvre et pleine de tics.

Des chercheurs canadiens montrent que les modèles spécialisés sont capables de comprendre la grammaire d'une langue. En revanche, les modèles généralistes, comme ChatGPT, sont incapables d'être réellement multilingues et ne transfèrent pas bien les règles grammaticales entre les langues. Connaître l'anglais n'aide pas vraiment une IA à parler le français québécois.

Les chercheurs indiens notent que l'IA efface les particularités de l'anglais indien pour produire un anglais standardisé. Les tics de langage entretiennent également un vaste système de déculturation. Ainsi, une description de Diwali (la fête des lumières) assistée par l'IA se concentre souvent sur des éléments occidentalisés, comme l'échange de cadeaux, tout en omettant les rituels religieux spécifiques et leur vocabulaire particulier. [...].

Fonte: Adaptado de : <https://theconversation.com/comment-de-ia-iser-nos-ecrits-pour-eviter-la-disparition-des-particularites-des-langues-281811>

6. **QUESTÃO 01 – Por que os textos produzidos por inteligência artificial são propensos a tornarem-se parecidos?** *

7. **QUESTÃO 02 – Por que o texto afirma que “Chacun écrit « mieux », mais tout le monde finit par écrire pareil.”?** *

8. **QUESTÃO 03 – O texto apresenta duas causas para os “tiques de linguagem” da inteligência artificial. Quais são elas?**

*

9. **QUESTÃO 04 – Qual é o risco das traduções automáticas produzidas pelas IA'S ?**

*

10. **QUESTÃO 05 –Segundo o texto, qual é um dos efeitos da inteligência artificial sobre o inglês indiano?**

*

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

