

Déchiffrer les scores des détecteurs d'IA : Le protocole de validation

Interpréter les scores de probabilité

- 0% - 20% : Probabilité faible, texte considéré comme majoritairement humain.
- 21% - 60% : Zone d'incertitude, le texte présente des structures hybrides ou trop formelles.
- 61% - 100% : Probabilité élevée, le texte présente des marqueurs syntaxiques typiques des modèles de langage.

Signes révélateurs d'une génération artificielle

- Répétitions syntaxiques : Utilisation récurrente de structures de phrases identiques.
- Manque de nuances : Absence de points de vue personnels ou d'anecdotes vécues.
- Structure trop parfaite : Absence d'erreurs humaines ou de variations de rythme naturel.
- Absence de sources récentes : Difficulté à citer des faits survenus après la date de coupure de l'IA.

Méthodologie de vérification humaine

- Fact-checking systématique : Vérifier chaque donnée chiffrée et chaque date citée.
- Analyse de la voix : Rechercher une signature stylistique propre à l'auteur.
- Test de cohérence logique : Identifier les contradictions subtiles ou les généralités creuses.
- Recherche de biais : Repérer les réponses trop neutres ou politiquement correctes.

Stratégies pour garantir l'authenticité

- Intégrer des expériences personnelles : Ajouter des détails contextuels impossibles à deviner pour une IA.
- Varier la longueur des phrases : Alternier entre phrases courtes percutantes et phrases complexes.
- Adopter un ton tranché : Exprimer des opinions fortes ou des points de vue singuliers.
- Citer des sources primaires : Référencer des documents ou des entretiens exclusifs.