# Un défi pour les détectives: repérer les traces de ChatGPT, Bard et autres robots conversationnels !

**La détection de cette pratique devient de plus en plus difficile avec les progrès de la technologie. Toutefois, franceinfo vous donne quelques astuces pour identifier ces textes qui ne sont pas rédigés par des humains.**

Avez-vous déjà imaginé un instant que l'article que vous avez lu en ligne ait été écrit par une machine ? C'est tout à fait possible. Grâce aux avancées fulgurantes de l'intelligence artificielle en matière de rédaction, il est désormais possible de rédiger des articles entiers à partir d'une simple instruction. Certains sites web publient ainsi plusieurs dizaines d'articles par jour, sans aucune supervision humaine, ou presque. Récemment, la start-up Newsguard, spécialisée dans la traque des sites de désinformation, a repéré 380 "sites d'actualité non fiables générés par l'IA". Il est donc tout à fait possible que vous ayez déjà lu l'un de ces articles sans vous en rendre compte, alors qu'ils sont très bien référencés sur les moteurs de recherche.

## ENQUÊTE. Comment l'intelligence artificielle s'invite à l'Assemblée nationale

Heureusement, une étude menée par une équipe de chercheurs de l'université de Pennsylvanie et publiée en décembre 2022 a montré que l'œil humain était capable, moyennant un peu d'entraînement, de reconnaître si un texte avait été rédigé par une personne ou par une IA. Pour mener cette étude, les

chercheurs américains ont conçu un jeu intitulé “Real or Fake Text ?”, dans lequel les participants devaient essayer de repérer les textes rédigés par un robot grâce à des indices subtils. Ce jeu, disponible en ligne, est uniquement en anglais, mais les mêmes principes peuvent être appliqués aux textes écrits en français. Voici quelques indices à prendre en compte lorsque vous lisez des articles en ligne pour éviter de vous faire berner.

1. Rechercher les répétitions (sans fautes)
   Un premier indice est que “les IA ne font pas de fautes d’orthographe”, explique Amélie Cordier, une ingénieure spécialisée en intelligence artificielle. Si vous trouvez une coquille (une faute de frappe, de grammaire, etc.) dans un texte, cela indique qu’il a été au minimum relu par un être humain. En revanche, les articles rédigés par une IA sans supervision humaine sont souvent truffés de répétitions. En effet, la génération de texte a tendance à reproduire les mêmes termes et structures de phrases, même si cela devient de moins en moins fréquent. Les IA sont de plus en plus performantes et leurs utilisateurs savent mieux les utiliser pour contourner ces problèmes. “Vous pouvez demander à l’IA d’écrire ‘à la manière de’, lui donner un exemple d’écriture, lui dire d’éviter les répétitions, et elle devient très performante et difficilement détectable”, explique Virginie Mathivet, une ingénieure et autrice d’une thèse sur l’intelligence artificielle à franceinfo. Des logiciels ont même été développés pour rendre les textes écrits par une IA encore plus humains. Le plus connu s’appelle Undetectable.ai et permet de “donner une apparence humaine” aux textes artificiels en les soumettant aux principaux outils de détection d’IA existants. Par conséquent, ces outils de détection deviennent de moins en moins fiables. “Open AI [l’entreprise créatrice de ChatGPT] a récemment abandonné son détecteur car ça ne fonctionne pas”,

souligne Virginie Mathivet.

2. Ils peuvent avancer des absurdités
   Les IA sont très performantes pour les tâches très codifiées, comme l'orthographe, mais elles peuvent aussi avancer des absurdités sans sourciller. "Une IA ne réfléchit pas, elle n'a pas de bon sens. Elle peut vous affirmer des choses absurdes avec une conviction absolue", note Amélie Cordier. "Si vous demandez à une IA d'écrire une recette d'omelette aux œufs de vache, elle peut tout à fait le faire", ajoute-t-elle. Les sites qui utilisent des IA pour produire des articles en masse à partir de contenus trouvés sur internet sont souvent confrontés à ce problème. Récemment, le site The Portal, qui traite de l'actualité du jeu vidéo, a été critiqué sur Twitter par le journaliste Grégory Rozières. Certains articles contenaient en effet de fausses informations évidentes, car l'IA qui les rédigeait avait pris au premier degré des blagues trouvées sur Reddit. Si vous lisez un article et qu'une information vous semble absurde, ou qu'un chiffre vous paraît démesuré, cela peut indiquer que l'article a été rédigé de manière non humaine. Pour s'en assurer, il est préférable de vérifier l'information douteuse auprès d'autres sources fiables. "Cela revient à vérifier les faits, c'est à l'humain d'adopter un regard critique", commente Virginie Mathivet.
3. Elles font preuve d'une productivité inhumaine
   La rédaction par IA ne garantit pas la qualité, mais elle permet de produire un grand nombre d'articles en un temps record. Soyez donc prudents face aux sites qui publient quotidiennement une quantité astronomique d'articles sans pour autant avoir un grand nombre d'employés. "Si vous constatez qu'un blog publie 200 articles par jour sous le même nom, cela peut être un indicateur", explique Virginie Mathivet. Certains articles rédigés par des robots portent la signature d'un nom, comme s'ils avaient été écrits par une

personne. Si cette signature semble trop fréquente, il y a de fortes chances qu'une IA soit utilisée. Sur le site The Portal, déjà mentionné précédemment, un même "journaliste" a signé près de 7000 articles en seulement neuf jours. De plus, si les articles présentent de nombreuses similitudes dans leur forme et leur structure, il est probable qu'ils soient rédigés automatiquement. En effet, les IA ont tendance à produire des contenus très homogènes, surtout lorsqu'ils sont créés à partir de la même instruction utilisée en boucle. "L'IA imite, c'est ainsi qu'elle fonctionne. Elle uniformise un peu tout", remarque Amélie Cordier.

4. Elles ont une écriture médiocre et citent rarement leurs sources
   Même si elles utilisent parfois des noms humains, les IA ne peuvent pas incarner leurs articles de la même manière qu'un journaliste réel. Si un journaliste n'a aucune existence en dehors de sa page auteur sur internet, cela peut être un indice laissant penser à une rédaction par IA. De plus, les articles publiés grâce à une intelligence artificielle ont tendance à manquer de cohérence et de citations de sources. Il est donc important de vérifier l'information à partir d'autres sources de confiance.

Source : www.francetvinfo.fr

## → Accéder à CHAT GPT en cliquant dessus

---

# Informatique secrète : Un examen final pas comme les autres s’est infiltré sur ChatGPT

Les limites de ChatGPT : une expérience éclairante

# Les limites de ChatGPT : une expérience éclairante

Dan Arena, professeur d’informatique à l’université Vanderbilt de Nashville (Tennessee), a voulu sensibiliser ses étudiants aux limites de ChatGPT. Son objectif était de les dissuader de l’utiliser pour tricher et de les encourager à développer leur esprit critique. Pour cela, il a posé des questions sur les algorithmes à ChatGPT, un agent conversationnel basé sur l’intelligence artificielle (IA). Certaines réponses étaient justes, tandis que d’autres étaient fausses, mais toutes étaient présentées avec tant de conviction qu’on aurait pu les croire authentiques, comme il l’a expliqué dans un témoignage publié par Slate.com. Cependant, Dan Arena n’avait pas prévu l’impact de cette expérience sur certains de ses étudiants.

Quelques jours avant la fin du semestre, l'un de ses étudiants est venu le voir et lui a fait part de son désarroi face à l'évolution de l'IA dans leur domaine : "Ces derniers temps, je suis déprimé par ce diplôme. Tout le monde parle de la façon dont les grands modèles de langage comme ChatGPT vont nous remplacer, nous les diplômés en informatique. J'ai l'impression que tout ce que j'ai appris au cours des quatre dernières années est déjà dépassé. Je ne sais pas quoi faire", a-t-il dit.

## Inception

Le professeur d'informatique lui a d'abord donné des conseils rassurants en expliquant qu'il devait apprendre à tirer parti de la technologie pour améliorer sa vie et sa productivité, plutôt que d'en avoir peur ou de supposer qu'elle ferait mieux que lui. Dan Arena lui a raconté que lorsqu'il a commencé à travailler, ses collègues utilisaient des cartes perforées pour coder et qu'internet n'existait pas encore. Et pourtant, il n'a pas encore été remplacé par une IA. Cependant, troublé par la détresse de l'étudiant, il a mis en place un plan secret pour dissiper ses inquiétudes : il a décidé de faire passer l'examen final de son cours sur les algorithmes à ChatGPT, en secret, pour voir s'il pouvait vraiment faire mieux que ses étudiants humains. Il a créé un faux profil d'étudiant nommé "Glenn Peter Thompson" et a attribué à ce dernier les résultats de ChatGPT, qu'il a intégrés dans la base de données des notes des élèves. Les résultats ont été surprenants.

"Chaque étudiant de mon cours du matin a obtenu de meilleurs résultats à l'examen final que Glenn, qui a obtenu un C- avec une note de 72,5%. La moyenne de ma classe se situait entre 80 et 90%. Quant à mon cours de l'après-midi, Glenn s'en est légèrement mieux tiré, mais il a tout de même obtenu des résultats inférieurs à la moyenne, dans le tiers inférieur de la classe, équivalent à un C+", révèle Dan Arena. Il précise

cependant que ces résultats ne sont pas définitifs, compte tenu des progrès rapides de la technologie. Cette expérience a rassuré les étudiants quant à leur avenir et a intéressé d’autres professeurs, qui envisagent de faire la même chose dans leurs cours.