

ChatGPT révélé : le détecteur ultime des textes écrits par l'IA!

Détecteur d'IA : comment distinguer les textes écrits par ChatGPT de ceux des humains



Comment distinguer les textes écrits par ChatGPT de ceux des humains ? Un détecteur y parvient, à 100%.

Mathilde Fontez, rédactrice en chef au magazine scientifique [Epsilon](#), nous parle aujourd'hui d'un détecteur impressionnant, qui déjoue les performances déjà extraordinaires de ChatGPT : comment distinguer des textes écrits par des intelligences artificielles, de ceux des humains ?

franceinfo : Des chercheurs américains viennent de mettre au point un détecteur qui parvient à distinguer des textes écrits par des hommes, et par l'IA ?

Mathilde Fontez : Oui, gros enjeu. On se souvient de la grève historique des scénaristes à Hollywood : il y a un mois, ils ont obtenu des garanties qu'un minimum d'humains continuera à écrire les scénarios. Mais encore faut-il distinguer ces textes des autres, mettre en place des contrôles...

Et le problème se pose aussi du côté de la science : la

plupart des revues scientifiques ont modifié leurs directives de publication pour interdire l'utilisation d'IA – parce que ChatGPT est bluffant, incroyablement performant, mais qu'il peut parfois "halluciner": écrire des choses totalement fausses, sans qu'on sache pourquoi, et ajouter des contrevérités dans des articles scientifiques.

C'est justement sur la science, que cette équipe de l'université du Kansas s'est concentrée. Elle a entraîné un programme à étudier des articles scientifiques dans des revues de chimie...

C'est une IA qui a été entraînée à détecter les IA ?

Eh oui, c'est l'IA qui contrôle l'IA – ça donne un peu le vertige. C'est un programme d'apprentissage, qui a été entraîné sur 100 articles écrits par l'humain, et 200 écrits par ChatGPT spécialement pour l'occasion. Les chercheurs lui ont donné des fonctionnalités pour examiner le style d'écriture : la variation de la longueur des phrases, la fréquence de certains mots, les signes de ponctuation...

Et il a réussi à détecter, dans 98 à 100% des cas, les textes qui avaient été écrits par ChatGPT. Il a même réussi avec des revues pour lesquelles il n'avait pas été entraîné.

C'est le premier détecteur de chatGPT ?

Non, il y en avait d'autres. OpenAI, l'entreprise qui a conçu ChatGPT, en avait déjà développé un. Un autre s'appelle zeroGPT. Mais leurs performances sont beaucoup moins impressionnantes : ils trouvent l'IA dans seulement 10 à 65% des cas.

Le point fort de ce nouveau détecteur, c'est qu'il est spécifique : il a été entraîné pour un type de texte en particulier : les publications scientifiques dans des revues de chimie. Et du coup, il est plus précis. Ça ne veut pas dire qu'il ne pourrait pas s'attaquer à d'autres textes, si on

modifie son entraînement.