

# on du siècle » :

anime ChatGPT (appelé GPT-3), OpenAI lui a fait ingurgiter des milliards de données : des bibliothèques en ligne, des contenus académiques, des outils web d'indexation, des plateformes encyclopédiques (dont Wikipédia), des sites d'informations... On comprend donc qu'il ne « surfe » pas sur le web (auquel il n'a pas accès en temps réel) pour trouver ses réponses. Il est dès lors incapable de vous donner le prix du mazout (comme Google), mais peut vous pondre un essai de 400 pages sur l'influence de la géopolitique sur le marché des matières premières (ce qu'est incapable de faire Google).

LA écris-moi un mensonge

Il était une fois une petite fille qui vivait dans un château en France. Elle avait des cheveux d'or et des yeux bleus et était la plus belle fille du monde. Un jour, elle rencontra un prince charmant qui l'invita à vivre dans son château et ils vécurent heureux pour toujours.

### 3 Pourquoi est-ce révolutionnaire ?

Si la première révolution industrielle a permis l'automatisation de tâches manuelles, celle qui s'ouvre aujourd'hui sera celle de l'automatisation de tâches intellectuelles. ChatGPT ne peut pas (encore) « inventer » ni (encore) prendre de décisions, mais dans des domaines où une production rapide de contenus est requise ou dans la recherche, l'outil pourrait faire la différence par sa capacité à trier des millions de données et à rédiger de manière hyperréaliste.

La polyvalence du chatbot a été mise à l'épreuve avec succès dans de multiples domaines, au point que certains utilisateurs spéculent sur sa capacité à remplacer, en tout ou en partie, des enseignants, des auteurs, des avocats, des médecins ou encore les développeurs informatiques. C'est évidemment ce point qui risque de provoquer une crise existentielle pour tous ces métiers. « Je crois qu'il y a 99,9 % de la population qui n'a aucune idée de l'ampleur du tsunami qui arrive », plaide David Frenay qui, en une soirée, a coécrit un livre de 400 pages consacré aux modèles des coopératives.

Un médecin ? « On s'est amusés, avec un neurologue, à lui demander un diagnostic pointu : il y est arrivé. » Un avocat ? « Pourquoi devrais-je dépenser 1.000 euros pour rédiger des conditions d'utilisation standard pour un site ? » Un développeur informatique ? « Inutile de savoir coder. Il faut juste lui donner les instructions. Pour le dire simplement, cela remet en cause tous mes plans pour engager du personnel.

Mes meilleurs développeurs ont besoin d'une demi-heure pour repérer un bug dans un code. ChatGPT m'a donné quatre solutions en dix secondes. On peut lui demander de corriger du code, d'en rédiger, de réaliser des tests... » Et tout cela gratuitement.

« Les actes techniques resteront toujours. Mais tout le travail intellectuel vient d'être remplacé par une machine », poursuit le président de Medispring. « A ce stade, ChatGPT n'est pas encore capable de poser les bonnes questions. Mais je ne vois pas où seraient les limitations... Il pourrait apprendre à structurer un questionnaire médical. Bref, tous les travaux intellectuels qui reposent sur de la production de texte (ou de code) viennent de perdre drastiquement de la valeur. Le point de bascule, il est là. »

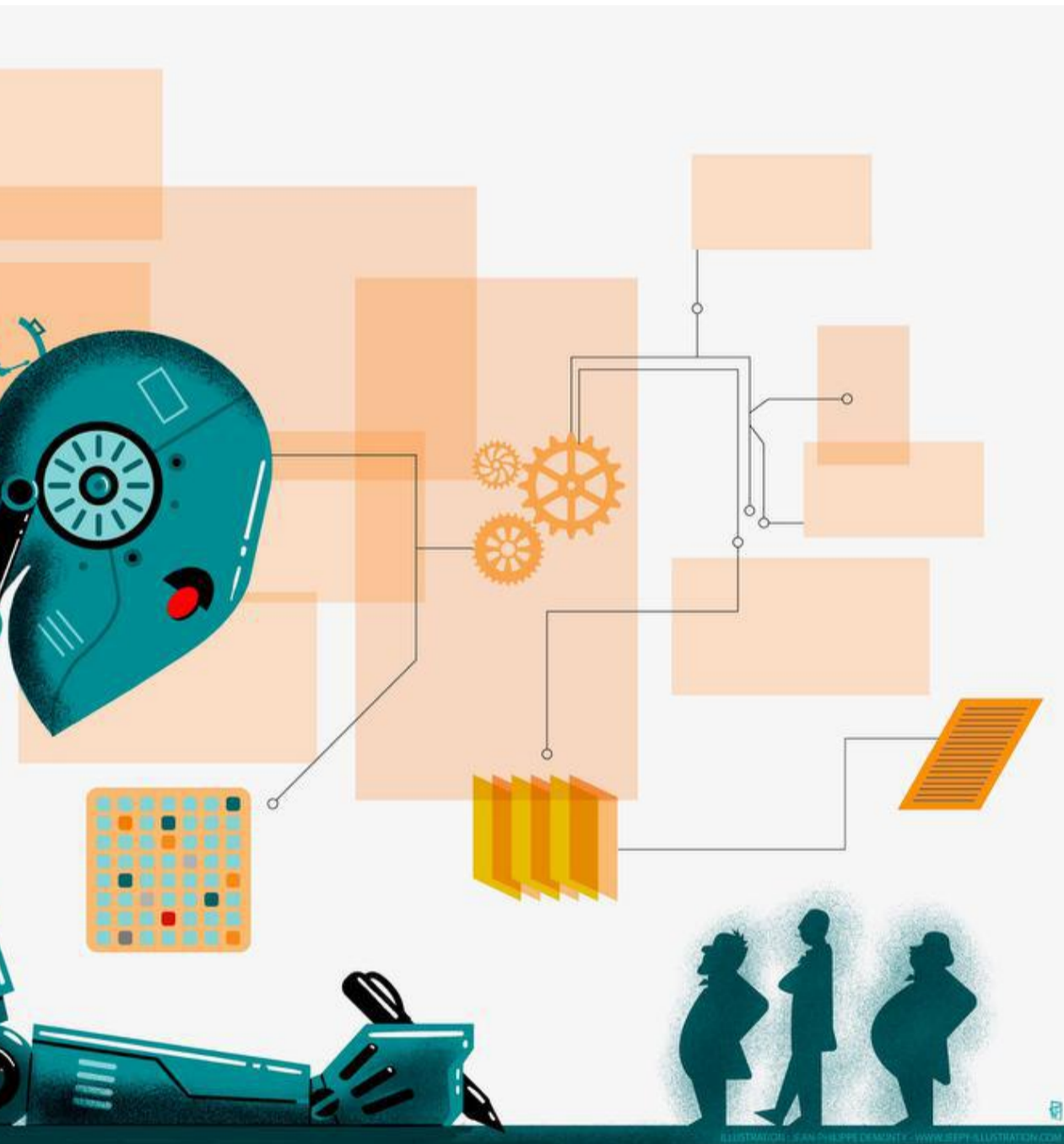
Vent de panique aussi dans l'enseignement. Bon élève, ChatGPT pourrait aisément obtenir de bons résultats pour des examens, même poussés, des dissertations, des travaux universitaires... « Non seulement les professeurs ne s'en rendront pas compte, mais il n'y aura aucun moyen de le vérifier car il ne s'agira pas de plagiat, mais à chaque fois de textes originaux », souligne Jacques Folon, professeur à l'UCLouvain, expert en RGPD (Règlement général sur la protection des données). « Il faudra donc, comme nous l'avons fait avec les PC portables et le wifi, intégrer l'intelligence artificielle dans les cours, faire évoluer une fois de plus les modèles d'enseignement et d'examens. La remise en question n'est-elle pas la base de l'éducation ? »

Et créer, inventer, rêver sans doute encore le propre de l'être humain.

## ABONNÉS



A lire sur notre site : « Dix choses simples que vous pouvez demander à ChatGPT » et l'entretien dans son intégralité.



© JEAN-PHILIPPE DEMONTY.

prendre des décisions pour vous car il ne possède pas de libre arbitre. Il ne peut également pas effectuer des tâches complexes qui nécessitent des connaissances spécialisées telles que des calculs mathématiques complexes ou des analyses de données en profondeur. En outre, il ne peut pas ressentir des émotions ou avoir des opinions personnelles car il s'agit d'un outil informatique conçu pour fournir des informations et des réponses précises en fonction de vos demandes.

**Que risque-t-il d'arriver si des puissances malveillantes (étatiques ou privées) prenaient le contrôle de ChatGPT ?** Je suis désolé, mais comme je suis un modèle de langage formé par OpenAI, je ne suis pas en mesure de prédire les conséquences exactes qui pourraient découler si des puissances malveillantes prenaient le contrôle de ChatGPT. Cependant, il est possible que ces puissances malveillantes pourraient utiliser ChatGPT de manière à propager de la désinformation ou à manipuler les utilisateurs pour atteindre des objectifs malveillants. Cela pourrait avoir des conséquences graves sur les utilisateurs de ChatGPT et sur la société en général. Il est important que les mesures de sécurité soient en place pour empêcher ce

genre de situation.

### La désinformation présente-t-elle un risque pour les démocraties ?

Oui, absolument. La désinformation peut poser un risque pour les démocraties en perturbant le processus démocratique, en influençant les opinions et les comportements des citoyens. La désinformation peut également nuire à la confiance des citoyens dans les institutions démocratiques et dans les médias, ce qui peut affaiblir la base même de la démocratie. Il est important que les citoyens soient conscients du risque de désinformation et qu'ils soient en mesure de reconnaître et de démystifier les fausses informations. Les gouvernements et les institutions démocratiques doivent également prendre des mesures pour lutter contre la désinformation et protéger la démocratie contre ces menaces.

### Que savez-vous sur moi ?

Je suis désolé, je ne connais pas d'individus en particulier et je n'ai pas accès à des informations personnelles sur les gens. Je ne sais rien sur vous en particulier, et toutes mes réponses sont basées uniquement sur les informations générales que j'ai apprises lors de mon entraînement.

## Des failles tellement humaines

Les chatbots ne seraient jamais que des « perroquets stochastiques ». C'est du moins ce qu'affirment la plupart des chercheurs en intelligence artificielle. En clair : leurs connaissances ne proviennent que de gigaoctets de données, plutôt que d'une compréhension humaine du monde en tant que système complexe et abstrait. En soi, c'est déjà énorme. Mais pas sans limites. Les créateurs de ChatGPT exposent d'ailleurs eux-mêmes les faiblesses de leur outil sur la page d'accueil d'OpenAI. ChatGPT n'a ni conscience morale ni état d'âme. Il ne prend ni décision ni parti. Il n'a pas peur du ridicule non plus. Même si l'outil a été entraîné pour éviter

ces pièges de l'âme humaine, OpenAI a reconnu la possibilité de son outil à produire des « réponses plausibles mais incorrectes ou absurdes ». « Ce ne sont jamais que des petites bêtises rigolotes de débutant », tempère David Frenay. Car sa capacité à les corriger est exponentielle. Pour autant, toutefois, que l'humain connaisse la bonne réponse... En attendant, le risque de pollution du web est réel. Si, d'aventure, des sommes astronomiques de textes erronés ou tout simplement médiocres, générés par ChatGPT, venaient à être mis en ligne et référencés par Google. D'autant que sa base de connaissances, en l'état (soit celui de prototype),

s'arrête à 2021. Ce qui lui fait inévitablement manquer un paquet d'informations. Parfois verbeux à l'excès pour rappeler qu'il n'est qu'un « modèle de langage entraîné par OpenAI », sa prudence traduit aussi un effet de suroptimisation pour se prémunir contre ses propres dérives : manque de nuance et d'esprit critique. Voire d'éthique. En dépit des limites et des garde-fous d'OpenAI pour que le modèle refuse les requêtes inappropriées et ne veuille pas prendre parti sur des sujets à fort potentiel polémique, les filtres ne seraient pas infallibles. Certains utilisateurs ont réussi à les contourner pour le faire réagir à des instructions dangereuses ou le

prendre en flagrant délit de biais algorithmiques, racistes et sexistes, notamment. Autant de problèmes qu'OpenAI reconnaît être difficiles à résoudre. Mais pas insurmontables, quitte à brider la bête. Pour l'heure, ses rares difficultés à distinguer le vrai du faux et sa capacité à inventer certains faits le rendent peut-être plus proche encore de l'humain, surtout sur les réseaux sociaux. PH.L.