DAILY SCIENCE
DÉCOUVREZ LA SCIENCE, LA RECHERCHE ET L'INNOVATION "MADE IN BELGIUM"

FACE AUX IA GÉNÉRATIVES, LES JOURNALISTES SONT PLUS QUE JAMAIS ESSENTIELS

Publié le 2 octobre 2023

par Laetitia Theunis

Série : JournaBot (1/6)

Nous vivons une révolution. Si l'invention de l'intelligence artificielle (IA) remonte à près d'un siècle, son usage dans le quotidien connaît un essor sans précédent depuis quelques années. La mutation digitale a peu à peu pris ses quartiers dans certains médias, confiant à une IA régie par des règles la rédaction d'articles de presse. Mais l'avènement de ChatGPT change radicalement la donne. Un nouveau paradigme d'IA a écrasé l'ancien modèle. Si l'on n'y prend pas garde, le caractère hallucinatoire des IA génératives ouvre grand la voie à la désinformation massive.

Code versus neurones artificiels

L'intelligence artificielle ne date pas d'hier. En 1943, l'ordinateur n'existait pas encore, mais sur papier, on concevait déjà le neurone artificiel. C'est l'évolution de cette vieille idée qui a abouti à ChatGPT. Ce système n'est piloté par aucune règle, il apprend tout par lui-même, et demeure encore aujourd'hui hors de contrôle des humains. Avec comme corollaire, une absence de garantie quant à la qualité et la véracité de ce qu'il produit.

Il a fallu attendre le tournant du deuxième millénaire, l'avènement de l'apprentissage profond (*deep learning*), l'accès à des quantités folles de données pour nourrir les réseaux de neurones et à

d'énormes puissances de calcul, pour que cette tradition neuronale s'impose. Et pousse sur le bas-côté celle qui a permis le développement de l'informatique automatisée durant près de 50 ans : la tradition symbolique.

« Celle-ci est directement inspirée du langage naturel, on parle d'ailleurs de NLP, pour *Natural Language Processing* ou traitement automatique du langage naturel. Le langage, qui est à la base de tout, a toujours été un Graal pour l'IA. Début des années 50, Alan Turing, un de ses concepteurs, considérait qu'une machine était intelligente si et seulement si elle nous confondait sur la manière de dialoguer », explique Pr Hugues Bersini, co-directeur d'IRIDIA, le laboratoire de recherche sur l'intelligence artificielle à l'ULB.

Pour ce faire, des règles syntaxiques, sémantiques, grammaticales extraites de la connaissance d'experts ou validées par ceux-ci, ont été codées dans les algorithmes. C'est ainsi que sont nés les premiers langages de programmation. « Le célèbre linguiste américain Noam Chomsky, l'un des penseurs de l'IA, a construit toute sa carrière sur la production de systèmes de NLP, notamment d'analyseurs syntaxiques. Si bien que dans les années 80-90, les premiers textes journalistiques ont été produits par une IA, de même que des textes de loi, des textes médicaux », poursuit Pr Bersini.

L'avantage des règles logiques, c'est que le contrôle est quasi absolu sur le contenu produit par l'IA. Et toutes les étapes de son processus de création sont comprises. C'est ce genre d'IA qui est utilisé depuis quelques années dans certains médias belges pour automatiser la rédaction d'écrits relayant les résultats des matches de foot locaux ou de la bourse.

Si elle paraît sécurisée, notamment dans le cadre strict du journalisme, cette tradition symbolique est aujourd'hui balayée par l'essor de la tradition neuronale, ou *machine learning*. Basée sur l'apprentissage automatique via l'analyse de quantités astronomiques de données, elle est, quant à elle, dénuée de sécurité.

Image libre de droit

Tsunami mondial

Sans conteste, il y aura un avant et un après ChatGPT. En novembre 2022, ce modèle neuronal d'intelligence artificielle était dévoilé par OpenAI. Il laissait la planète bouche bée par sa capacité à

répondre à des questions, à écrire des textes argumentés et des poèmes, à générer du code informatique, à traduire des langues. GPT-4, la dernière version, datant de mars 2023, s'attaque aux images : elle est capable de les interpréter, de les analyser, d'en faire une description précise et de suggérer de nouvelles idées sur cette base.

A noter que le paysage des IA génératives, comme on les appelle désormais, ne se limite pas à ChatGPT. En janvier 2023, Microsoft, principal investisseur et partenaire d'OpenAI, ajoutait un *chatbot* similaire à Bing, son moteur de recherche Internet. Un mois plus tard, LlaMA, de Meta, poussait son premier cri. En mars 2023, naissaient Bard, le *chatbot* de Google et Ernie, premier rival chinois de l'américain ChatGPT, développé par Baidu.

Nombreux parmi les auteurs et les journalistes ont alors craint d'être remplacés par des algorithmes à court ou moyen terme. Par ailleurs, l'usage de ces IA par des quidams dans le but de générer des textes colportant des informations fausses, notamment pour détourner l'attention du public dans le cadre d'élections, fait pressentir l'avènement d'une ère de désinformation massive.

En effet, ChatGPT et ses cousins, s'ils savent écrire des textes semblables à des articles de presse, n'assurent en rien la véracité des faits qu'ils évoquent dans leurs lignes : ils en sont bien incapables.

Un baratineur de génie

Tous fonctionnent sur le même principe. Si ChatGPT est capable d'écrire un texte cohérent, c'est parce qu'il a été poussé à se construire tout seul une représentation de la façon dont fonctionne le langage. On parle de *machine learning* et de *deep learning*. Pour ce faire, il a avalé des quantités phénoménales de textes écrits dans les langues dominantes que compte actuellement la planète, mais sans jamais en comprendre le sens.

« Il a lu l'entièreté de Wikipédia. Et cela ne représente que 3 % de toutes ses lectures. Il s'est aussi nourri de livres et d'articles de presse en s'asseyant allègrement sur la notion de copyright ainsi que de forums Internet avec tous les risques que cela comporte », précise Benoît Frénay, professeur en apprentissage automatique (*machine learning*) au sein de l'institut NADI de l'UNamur. Depuis novembre 2022, de nombreuses plaintes ont été déposées en justice contre les entreprises productrices d'IA génératives par des auteurs. Et ce, pour avoir utilisé le contenu de leurs œuvres sans leur consentement ni octroi d'une rémunération.

« ChatGPT, c'est une grosse boîte noire appelée réseau de neurones artificiels, qui apprend à construire le mot suivant. Pour cela, il se crée des représentations internes. Pour prendre une analogie, un modèle similaire en traitement de l'image apprend à reconnaître des roues sur des photos. ChatGPT, lui, apprend à reconnaître des structures dans les textes. » Autrement dit, ses puissants réseaux de neurones étant capables de distinguer des ressemblances dans les textes écrits par les humains qu'il a lus (par exemple, qu'une histoire de roi est suivie d'une histoire de guerre avec un haut pourcentage statistique), il produit des textes dans lesquels ces similarités sont présentes. « ChatGPT agit comme l'autocomplétion qui existe sur votre GSM et qui propose le mot suivant à ce que vous êtes en train de taper. C'est uniquement cela. »

Que se passe-t-il lorsqu'il reçoit une demande du type « écris-moi un article critique d'une page sur la musique arabe au 12e siècle » ? Il pond simplement la suite probable. « Ce qu'il essaie de faire, c'est d'écrire un texte ni vrai ni faux, mais juste vraisemblable. C'est un baratineur de génie. C'est un *bullshiteur*. C'est le collègue qui a son avis sur tout et n'importe quoi, et que l'on n'arrive jamais à coincer», poursuit Pr Frénay.

Dénué d'intelligence et de sens commun

ChatGPT ayant été entraîné à rédiger des textes vraisemblables, ce qu'il écrit est de prime abord bluffant, extraordinairement ressemblant à ce qu'un humain produirait.

« Si on lui demande d'écrire une lettre de rappel pour non-paiement de facture, il excelle. Ce genre

d'écrit, c'est uniquement du style, dénué de factuel. Mais quand on lui demande un écrit basé sur des faits, il est capable d'inventer n'importe quoi. »

Le Pr Benoît Frénay s'est amusé à lui demander qui a tué Albert 1er. Pour rappel, alpiniste, l'ancien roi des Belges est mort à la suite d'une chute accidentelle des falaises de Marche-les-Dames alors qu'il s'entraînait pour de futures ascensions alpines. En plus de cette version officielle, ChatGPT révèle qu'à l'époque, il y a eu d'autres théories, notamment d'assassinat, et les explique. « Quand je lui demande d'où viennent ses sources, il en invente des vraisemblables, mais qui n'existent pas : un reportage de la RTBF, un livre d'un vrai auteur qui a écrit sur des sujets similaires. Bien que ce livre n'ait jamais été écrit, ChatGPT est capable d'en résumer un chapitre… Il n'est pas conscient, il n'a pas de mémoire, il n'est pas intelligent, il ne sait pas que ce livre n'existe pas. C'est juste une grosse équation, un gros modèle mathématique de langage. On parle de *large language model* (LLM). »

C'est bien cela qui est dangereux: ChatGPT est capable d'écrire des textes qui ressemblent à de véritables articles émanant d'un travail journalistique rigoureux, mais risque d'inventer des faits tout en sourçant erronément. Avec, à la clé, une désinformation massive.

Image libre de droit

Garde-fous nécessaires

En janvier 2023, un scandale éclatait. Une enquête du magazine américain Time révélait qu'OpenAI avait recours à des Kényans sous-payés (moins de 2 euros par heure) pour filtrer les contenus toxiques de ChatGPT. En effet, celui-ci étant notamment alimenté par des contenus du web, il présentait de très forts biais sexistes, racistes et violents. Pour tenter de lui ôter cette tendance, les travailleurs africains devaient lire des textes contenant une litanie d'horreurs et les étiqueter. Ceux-ci étaient ensuite soumis à l'IA générative comme exemples de ce qu'elle ne devait pas faire. « Le plus important, c'est de se rendre compte que ChatGPT (et ses cousins, NDLR) peut partir en roue libre si on n'y prend pas garde », alerte Pr Frénay.

« Ce processus de nettoyage des données est assez complexe. C'est la raison pour laquelle ChatGPT a arrêté ses lectures en septembre 2021. Par ailleurs, le langage n'ayant pas évolué en 2

ans, pas besoin de mettre le système à jour. »

Au vu de ses écueils, l'usage des IA génératives généralistes devrait s'arrêter, en journalisme, à trouver un titre attrayant, condenser un texte, enlever des lourdeurs, utiliser un ton moins formel, améliorer un texte via des reformulations, en changer le style, le condenser ou encore restructurer des notes. Mais attention, sans jamais faire une confiance aveugle : la relecture est primordiale.

Genesis se dédie au journalisme

En parallèle, Google teste Genesis, un outil d'intelligence artificielle qui interviendrait dans la rédaction d'articles de presse. Il ne s'agit plus d'une IA généraliste comme ChatGPT, Bard ou Ernie, mais d'une IA générative spécialisée dans le journalisme. Autrement dit, elle serait entraînée sur un corpus composé essentiellement d'articles de presse, de différents styles, de différentes lignes éditoriales. Le développement de l'outil et ses premiers essais ont débuté en juillet 2023 avec le Washington Post, le New York Times et le Wall Street Journal.

Peu d'informations filtrent sur ce projet. Selon un communiqué de Google, il s'agirait d'une aide à l'écriture d'articles, proposant différents titres et styles d'écriture. En automatisant certaines tâches, Genesis servirait d'assistant personnel aux journalistes. Leur libérant ainsi du temps pour s'adonner au reportage de terrain, à la vérification et au recoupement des informations, à leur approfondissement et à leur analyse. Toutefois, les limites potentielles du futur outil sont floues. Peut-être Genesis ira-t-il jusqu'à rédiger des articles au départ de notes en vrac, voire directement sur base de flux de données ?

Rêve ou cauchemar ? « Dès lors que l'outil est conçu par une plateforme extérieure, le vrai risque, c'est que les médias, les rédactions, n'aient pas la pleine maîtrise à la fois de son fonctionnement et de son produit final. Et s'ils n'ont pas la maîtrise de la boîte noire, cela veut dire qu'ils délèguent l'intervention sur les contenus à quelque chose d'extérieur, mettant à mal la notion même d'indépendance journalistique», commente Muriel Hanot, secrétaire générale du conseil de déontologie journalistique (CDJ).

Dans le cas où des rédactions envisageraient un modèle intégré d'IA comme aide à la production, la recommandation déontologique est donc d'en avoir la pleine maîtrise. « C'est un des fondements de la déontologie du journalisme », conclut-elle.

Cette large enquête sur le journalisme automatisé par l'intelligence artificielle a bénéficié du soutien du Fonds pour le journalisme en Fédération Wallonie-Bruxelles.

JournaBot 1/6 : Face aux IA génératives, les journalistes sont plus que jamais essentiels

JournaBot 2/6 : Quand les médias rejoignent le portefeuille client d'entreprises informatiques

JournaBot 3/6 : À l'heure de l'IA, les journalistes doivent jouer leur rôle de certificateurs de l'information

JournaBot 4/6 : Un cadre juridique pour freiner d'urgence la folle accélération de la désinformation

JournaBot 5/6 : La nature du travail journalistique en pleine mutation

JournaBot 6/6: Une avalanche d'outils à découvrir, à maîtriser, et (peut-être) à utiliser