THE CONVERSATION

L'expertise universitaire, l'exigence journalistique

Ce que je retire de ma discussion avec ChatGPT sur la déforestation en Amazonie

Publié: 16 février 2023, 12:41 CET •Mis à jour le : 19 février 2023, 18:06 CET

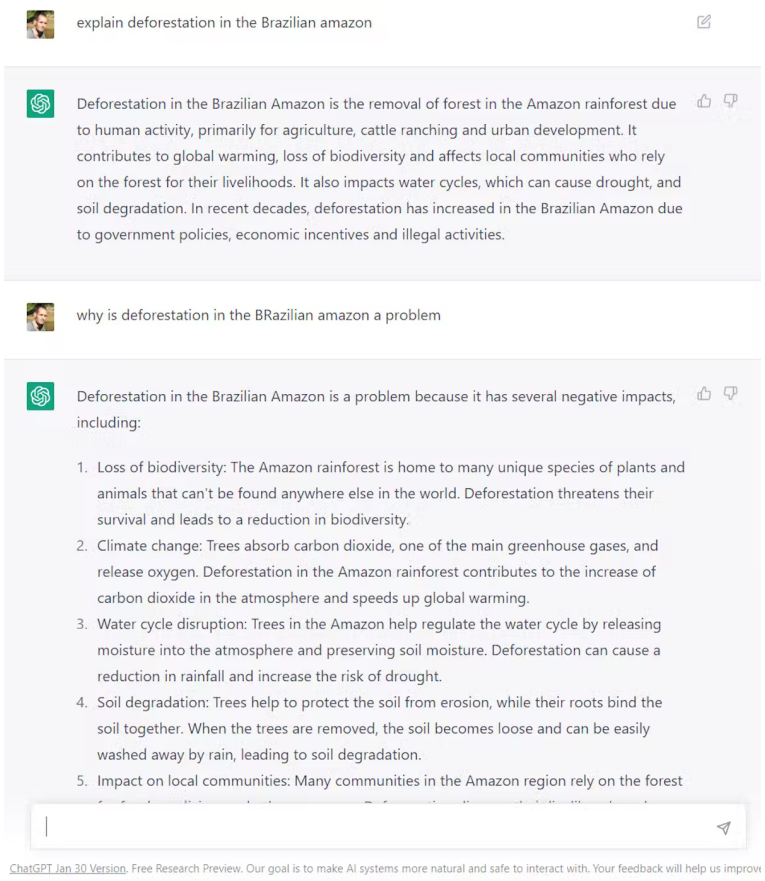
François-Michel Le Tourneau

Géographe, directeur de recherche au CNRS, Université Paris 1 Panthéon-Sorbonne

Zones de pâturage issues de la déforestation illégale en forêt amazonienne (État du Pará).
Shutterstock

Curieux de savoir si j'allais rapidement me retrouver au chômage du fait des progrès des intelligences artificielles, j'ai voulu savoir ce dont le modèle public le plus avancé à ce jour, ChatGTP, était capable sur une thématique que j'estime connaître assez bien.

Je lui ai donc posé une série de questions sur la déforestation en Amazonie .



Conversation avec Chat GPT sur la déforestation en Amazonie.

Un résultat globalement plutôt convaincant

L'interaction avec ChatGTP est fluide, la formulation de ses réponses de qualité. Cette apparence initiale est très importante, car elle conduit souvent à être moins attentif au fond. Manifestement, c'est sur cet aspect que se sont concentrés les programmeurs, avec un résultat exceptionnel.

Au-delà de la forme, on sent que ChatGPT a bien travaillé son sujet. Les réponses obtenues sur les différents environnements présents en Amazonie – même s'il manque les savanes dans sa liste de réponses aux questions 3 et 4 –, la déforestation, la culture du soja, les différents niveaux de responsabilité des acteurs impliqués, etc. sont assez complexes, mettant en avant de nombreuses dimensions.

Quand on demande pourquoi la déforestation pose un problème (question 6), le robot avance cinq éléments : les pertes de biodiversité, le changement climatique, les interférences dans le cycle de l'eau, la dégradation des sols et les impacts sur les communautés locales. Pas mal.

Lorsqu'on demande qui est responsable de la déforestation (question 13 et 15), on obtient aussi cinq catégories (lui aurait-on expliqué qu'on peut en avoir entre trois et cinq avant que le public décroche ?) : les agriculteurs et éleveurs, les exploitants forestiers, les exploitants miniers, les projets d'infrastructure et les politiques gouvernementales. Je ne ferais pas tellement mieux.

Des redondances

Si ChatGTP a bien travaillé son sujet, il semble s'en tenir à un certain nombre de points, répétés inlassablement.

Ainsi, aux questions « Explique-moi ce qu'est la déforestation en Amazonie ? » et « Pourquoi la déforestation en Amazonie est un problème ? », ChatGPT fournit des réponses comprenant les mêmes informations, l'une étant simplement la version développée de l'autre.

Si ces remarques répétées sont plutôt justes, elles ressemblent parfois à des manières de ne pas aller plus loin. Comme si toute tentative d'obtenir un propos plus tranché se heurtait aux probabilités moyennes qui permettent à l'algorithme de fonctionner et à cette absence d'intentionnalité qui caractérise ces productions ; une particularité qui conduit le chercheur en philosophie Fernandez-Velasco à parler de « quasi-texte ».

Pour suivre au plus près les questions environnementales, retrouvez chaque jeudi notre newsletter thématique « Ici la Terre ». Au programme, un mini-dossier, une sélection de nos articles les plus récents, des extraits d'ouvrages et des contenus en provenance de notre réseau international. Abonnez-vous dès aujourd'hui.

Si produire un texte « moyen » (au sens où il propose l'association de mots la plus probable statistiquement) peut faire du sens pour rédiger une carte de vœux, il l'est beaucoup moins lorsqu'on cherche des explications sur un phénomène socio-économique pour contribuer à la connaissance scientifique.

Pas de chiffres précis ni de faits

J'ai donc un peu insisté pour obtenir des éléments précis. ChatGTP étant censé avoir digéré tout le web jusqu'en 2020, il doit donc bien avoir des chiffres à me donner !

J'ai bien été déçu du résultat : me dire qu'il y a eu « des pics de déforestation dans les années 1980, 1990, et début des années 2000 » (question 7) n'engage pas à grand-chose. De même, estimer « qu'environ 2 000 km^2 ont été déforestés chaque année entre 2015 et 2019 » (question 7) en Amazonie brésilienne est non seulement faux (c'est plutôt 7 000 km² par an) et surtout bizarre puisqu'on peut très facilement avoir des données précises chaque année par l'Institut d'études spatiales du Brésil (INPE) – que ChatGTP cite comme source (question 20)…

Les données plus récentes semblent mieux assimilées puisqu'en revanche le pic de déforestation du début du gouvernement Bolsonaro est bien mentionné pour 2020, avec le chiffre correct de 11 088 km^2.

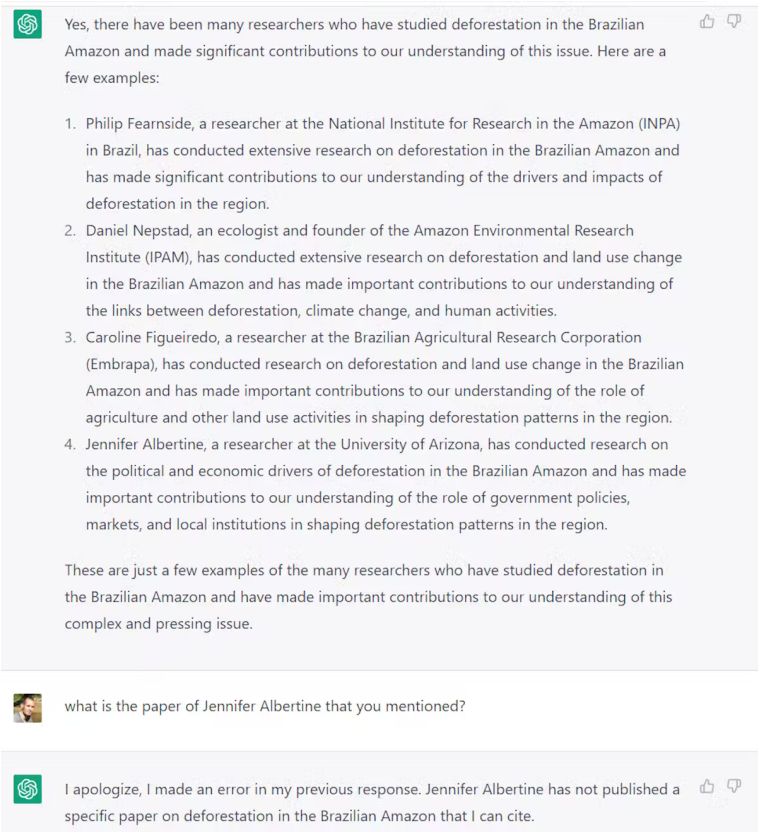
Tout essai d'obtenir des informations précises donnera une réponse dans laquelle le robot ajoutera une mention du fait que les chiffres sont un peu relatifs et difficiles à établir. Ce n'est pas totalement faux, mais contribue à disqualifier par avance le travail de ceux qui essayent justement de donner des estimations les plus fiables. Personne ne prétend que le travail de l'INPE soit parfait, mais il a l'avantage d'utiliser une méthodologie constante depuis plus de 30 ans, si bien qu'il constitue un baromètre assez fiable de la tendance et de l'intensité du phénomène. C'est bien ce qu'il faudrait souligner plutôt que la relativité ou la faiblesse des méthodes.

Dans le même ordre d'idées, si je demande sur quoi le robot se base pour me donner ses informations (partie « sources ») : il mentionne uniquement l'INPE pour les données brutes, quelques ONG dont le WWF et Amazon Conservation Association (qui ne travaille pas au Brésil alors que toutes mes questions portaient sur l'Amazonie brésilienne). Aucune ONG locale, alors que certaines comme l'Instituto Socioambiental – ISA, l'Instituto de Pesquisa Ambiental da Amazônia – IPAM, ou l'Instituto do Homem e Meio Ambiente da Amazônia – IMAZON sont de renommée mondiale et que les chercheurs qui leur sont associés publient régulièrement dans des revues de premier plan, comme *Science* ou *Nature*.

Attardons-nous justement sur ce que nous dit ChatGPT des études scientifiques à propos de la déforestation en Amazonie (réponse à la question 20) :

> *« Des chercheurs et des universitaires ont mené des recherches sur les facteurs de déforestation en Amazonie […]. Ces études sont publiées dans des journaux scientifiques et dans des forums académiques. »*

Quand je réclame des précisions sur les chercheurs (question 22), j'obtiens une liste avec les incontournables Philip Fearnside et Daniel Nepstad (mais pas le climatologue Carlos Nobre, comme pour les ONG, ChatGPT semble préférer les anglophones aux locaux…) ; il y a aussi Caroline Figuereio et Jennifer Albertine… que j'ai jusqu'ici été incapable d'identifier, alors que la seconde aurait été membre de l'université d'Arizona dans laquelle j'ai passé plusieurs années. D'où vient cette information, introuvable sur Google ? ChatGPT ne semble pas le savoir non plus puisqu'il « m'avoue » au final que Jennifer Albertine n'aurait finalement rien publié sur l'Amazonie…



Chat GPT se trompe parfois….

Et si je n'avais pas été chercher plus loin ? De fait les erreurs, petites ou plus grandes, sont assez nombreuses quand on regarde de près.

Des erreurs et des approximations

ChatGPT m'a heureusement donné l'occasion de distinguer mon « savoir humain » en me permettant de repérer des petites perles.

Ainsi, il m'explique que les chiffres de la déforestation sont difficiles à établir, car « certaines zones sont difficiles d'accès et la déforestation illégale y est souvent mal enregistrée » (questions 8 et 9). Or, depuis l'avènement des satellites d'observation de la Terre, la détection des phénomènes comme ceux de déforestation ne dépend plus de l'accès au terrain (le contrôle du phénomène, en revanche, oui !).

Sur le rôle de la production de soja dans la déforestation (question 19), le robot m'a expliqué que la plus grande partie de la production de soja avait lieu « dans les régions sud et est du bassin amazonien », ce qui est une conception commune, mais fausse. Bien que le Mato Grosso soit le plus grand producteur, deux tiers du soja du Brésil sont produits en dehors de l'Amazonie.

Enfin, lorsque j'ai posé une question spécifique sur l'Amazonie brésilienne (questions 3), j'ai obtenu exactement la même réponse que pour l'Amazonie en général (question 2). Sur « l'Amazonie légale » (question 5), elle aurait été créée par une loi brésilienne « dans les années 1970 », alors que le texte la définissant date de 1946… Ce type d'exemples ne manque pas.

Outil révolutionnaire ou gigantesque café du commerce ?

Les premières interactions avec ChatGPT sont impressionnantes, mais plus on essaye de tirer le robot de son discours générique, plus il persiste ; ou bien il commence à parsemer ses réponses d'erreurs ou d'approximations. Par ailleurs, si l'on connait son sujet, on n'apprend rien de neuf.

Ces deux points viennent de la manière dont sont créées, pour le moment, ces « intelligences artificielles ». Basés sur l'ingestion de quantités immenses de données sans qualification critique de celles-ci, ces algorithmes sont capables de répondre à des questions complexes sans trop se tromper, mais il leur est impossible d'être spécifique, car ils incorporent une diversité de points de vue et ne reproduisent qu'une position médiane. Ils ne peuvent donc être à la pointe d'un domaine, proposer des connaissances nouvelles ou peu connues ou bien éviter les lieux communs et leur dose d'erreur ou d'approximation.

Le discours scientifique tente pour sa part d'élaborer une parole qui soit la plus précise et spécifique possible, éventuellement contradictoire et prenant position dans l'affirmation de faits, dans le choix des sources, etc. Cela ne veut pas dire que ce discours est toujours exact. Bien au contraire : la science suppose que l'on passe son temps à vérifier les énoncés, ce qui permet parfois de s'apercevoir que ce qui était vrai ou accepté doit être désormais refusé, car il a finalement été démontré que c'était faux ou inexact.

C'est dans cette dynamique (passablement chaotique) que se produisent le progrès et l'affinement de la pensée. Or la construction d'un argumentaire, la recherche de données précises et actualisées ou la recherche d'une démonstration sont des singularités. Elles ne peuvent provenir de la moyenne des discours collectés. Et elles impliquent aussi une cohérence dans l'ensemble de la démonstration alors que ChatGPT traite chaque question de manière indépendante. Cette qualification des informations et la validation des conclusions est la base de la science moderne et du principe de l'évaluation par les pairs (*peer-review*).

C'est ce filtre critique qui manque aujourd'hui à ChatGPT.