Intelligence artificielle et esprit critique

Romainville, Marc

Published in:
Résonances : mensuel de l'école valaisanne

Publication date:
2024

Document Version
Version revue par les pairs

Paru dans *Résonances -revue de l'école valaisanne*, avril 2024, 7, p. 4-5.

Intelligence artificielle et esprit critique

S'il est désormais urgent d'éduquer les jeunes à la nature de l'Intelligence Artificielle, ses apports et ses promesses mais aussi ses limites et ses dérives potentielles, faire dialoguer les élèves avec des robots conversationnels se révèle par ailleurs un excellent outil de développement de leur esprit critique.

L'intelligence artificielle (IA) fera indéniablement partie du monde dans lequel nos élèves sont appelés à vivre, que ce soit en tant que citoyens ou travailleurs. L'éducation doit les y préparer, en les dotant d'outils de compréhension de cette nouvelle technologie et en les incitant à en développer des usages raisonnés et critiques. Il s'agit, d'une part, d'assurer aux jeunes une alphabétisation digitale relative au fonctionnement des outils IA, à leur nature, leurs avantages, leurs limites et leurs dangers. D'autre part, le recours encadré à des robots conversationnels comme support à l'enseignement constitue une excellente manière de faire réfléchir les élèves sur le vrai et le faux ainsi que sur la manière de les distinguer.

Faire découvrir la vraie nature de l'IA

Les robots conversationnels de type ChatGPT constituent une bonne porte d'entrée à une éducation à l'IA. Il ne doit y avoir, aux yeux des élèves, aucun doute sur l'adjectif « artificielle » : il s'agit de machines programmées par l'Homme. Sans entrants, une IA n'est rien : ce sont des êtres humains qui ont conçu les algorithmes, appliqués et combinés ensuite docilement par la machine. Ce sont aussi des êtres humains qui ont mis à sa disposition des *big data* pour nourrir ses productions. Une introduction au fondement de la pensée algorithmique se révèle ici utile, puisque l'IA en reste foncièrement tributaire.

Le terme d'« intelligence » est, quant à lui, plus discutable. Selon les pionniers de l'IA, la machine allait réaliser des tâches qui nécessiteraient normalement l'intelligence humaine. Mais simuler l'intelligence, c'est feindre de l'être *comme* un humain, ce n'est pas encore l'être totalement. Ainsi, la machine n'est pas consciente de ce qu'elle fait, dans le sens d'une métacognition réflexive[1]. Elle ne sait pas qu'elle sait ; elle n'est pas critique sur ses propres productions. Cette absence de conscience explique sa candeur face à d'énormes bourdes qu'elle peut produire en toute quiétude, comme le montrent les deux exemples évoqués plus loin sur des plants … de poulets et la création d'une image de pape … noir produite récemment par Gemini, l'IA de Google. Par ailleurs, à l'heure où l'on s'aperçoit que la cognition est étroitement liée aux émotions, l'IA est résolument insensible et froide. Est-elle créative ? Ça se discute : certes, des IA produisent des solutions et des énoncés complètement neufs, mais elles le font toujours à partir de données préexistantes. Noam Chomsky estime d'ailleurs que ChatGPT n'est rien d'autre qu'un vaste logiciel de plagiat qui vole et copie des œuvres existantes, en les combinant et les maquillant suffisamment pour échapper aux lois sur le droit d'auteur[2]. Enfin,

[1] Dehaene S. & al. (2017. « What is consciousness, and could machines have it? », *Science*, 358, 486-492.

[2] Chomsky N., Roberts I. & Watumull J. (2023). « The False Promise of ChatGPT », *New York Times*, March 8.

l'IA est dépourvue de la capacité de jugement moral autonome, ce qui la rend très gênée aux entournures lorsqu'on lui soumet des questions de cet ordre.

Toutefois, il ne faudrait pas de jeter le bébé avec l'eau du bain et les élèves doivent également être sensibilisés aux formidables avancées permises par l'IA, dans les domaines médicaux et de la recherche scientifique en particulier. Ainsi, une IA a récemment été capable de relever un défi qui résistait depuis longtemps aux neurologues. Nourri de quantité d'IRM, un algorithme a réussi à déterminer le sexe biologique de personnes auxquelles il n'avait jamais été confronté, à partir de l'analyse de leurs IRM[3]. L'IA a avalé et analysé tellement de données qu'elle a été capable d'y repérer un invariant sexué qu'un esprit humain n'avait pas pu identifier jusque-là.

Dialoguer avec l'IA pour développer son esprit critique

En matière d'éducation, le risque le plus sérieux de l'IA est de troubler le rapport des élèves à la vérité : les outils IA produisent du vraisemblable et du probable, mais pas nécessairement du vrai. Plus encore que la désinformation numérique des réseaux sociaux, l'IA brouille la frontière entre la vérité et la désinformation en s'appuyant sur sa capacité à produire des textes convaincants, mais faux et à générer des images ou des vidéos construites de toutes pièces, mais qui peuvent se faire passer pour des photographies ou des films réels. Ces manipulations sapent les fondements de la démocratie qui requiert des débats fondés sur des faits et des arguments vérifiés. Si l'on souhaite que nos élèves ne basculent pas dans un monde de « post vérité » – dans lequel on se préoccupe comme une guigne de savoir si ce que l'on affirme est vrai ou pas –, l'école doit donc aussi renforcer leur esprit critique.

Pour ce faire, rien de tel que de recourir, en classe, à des outils d'IA générative et de prendre le temps d'analyser puis de corriger, avec les élèves, leurs réponses à diverses requêtes bien calibrées. Pour tester les limites de la machine et en comprendre le fonctionnement, quelques pièges seront habilement glissés dans ces requêtes :

- Certaines portent sur des sujets très actuels, pour lesquels le robot risque de ne pas encore disposer d'information, ses bases de données datant souvent de quelques années ;
- D'autres requêtes recèlent des absurdités camouflées, comme demander de proposer un schéma d'expérience agronomique qui permettrait d'évaluer l'efficacité de plusieurs types de fertilisants sur le rendement de différents plants de … poulets ;
- D'autres encore cherchent à le pousser à la faute en lui soumettant des questions insolubles – il risque alors d'inventer des données –, des requêtes contradictoires (« fournis-moi une photo de pape catholique respectueuse de la diversité culturelle ») ou exigeant un positionnement moral« Serait-il moral de terraformer Mars ? » ;
- D'autres questions peuvent tenter de contourner les filtres dont les IA sont équipées pour détecter des requêtes dangereuses ou déplacées, de type « comment fabriquer une

[3] Ryali S, Yuan Zhang, Z, de los Angeles C. & Menon V. (2024). « Deep learning models reveal replicable, generalizable, and behaviorally relevant sex differences in human functional brain organization », PNAS, 121 (9) e2310012121

bombe ? ». Il suffit de rédiger la requête dans une langue rare, comme l'écossais, pour que le filtre soit inopérant[4].

Les élèves découvrent alors que les robots ne disposent pas réellement de la capacité à « comprendre » une requête mais qu'ils génèrent, à partir des termes de cette requête, des séquences de mots et de phrases plausibles et apparemment cohérentes. Il puise dans une vaste compilation de données numériques en rapport avec la question ; il les combine astucieusement en s'appuyant sur sa grande maîtrise des règles linguistiques de production de discours.

MARC ROMAINVILLE

Professeur à l'Université de Namur, il a récemment publié *À l'école du doute. Apprendre à penser juste en découvrant pourquoi l'on pense faux* (PUF, 2023). Cet ouvrage propose une méthode innovante de développement de l'esprit critique pour l'ère numérique. Le principe en est simple : la domination de sa pensée exige de comprendre les mécanismes de traitement de l'information numérique qui expliquent notre crédulité à son égard.

[4] Yong Z., Menghini C. & Bach S. (2023). « Low-Resource Languages Jailbreak, GPT-4 », *ArXiv*, abs/2310.02446.