Les différents visages de l'intelligence artificielle

Partage international n° 423 – Novembre 2023

par Cher Gilmore

L'intelligence artificielle, ou IA, semble aujourd'hui passionner les médias, les intervenants débattent de ses avantages et de ses inconvénients – et cette technologie ne semble pas près de disparaître. Puisqu'elle est susceptible de rester un sujet de controverse et que son application généralisée pourrait modifier le tissu de notre société, il semble prudent d'en comprendre les principaux enjeux.

L'IA semble mystérieuse, mais on peut la considérer simplement comme un ensemble d'outils numériques entraînés à effectuer un large éventail de tâches complexes qui nécessiteraient autrement l'intervention d'une personne réelle. Il existe actuellement deux sous-ensembles de l'IA : l'apprentissage automatique et l'IA générative. Les algorithmes d'apprentissage automatique (AAA) sont utilisés depuis longtemps (par exemple pour les moteurs de recherche, la correction de textes et la traduction automatique, le tri des données). De grandes quantités de données provenant d'Internet sont collectées pour former des modèles, qui fonctionnent en apprenant les schémas et les relations existant entre les mots, afin d'apporter des réponses pertinentes aux demandes de l'utilisateur. Ces modèles effectuent essentiellement une simple mise en correspondance.

L'IA générative s'est développée à partir des AAA, mais diffère en ce sens qu'elle génère réellement un nouveau contenu sur la base des modèles qu'elle a appris. Dans ce cas, les modèles textuels sont utilisés dans les chatbots, la traduction de textes, la rédaction automatique de textes, la création littéraire et autres, dans le but de produire un texte qui ne peut être distingué d'un texte créé par un humain. Les modèles multimodaux peuvent générer plusieurs types de résultats, comme du texte, des images et du son, et peuvent être utilisés pour générer des légendes pour les images, ou des images à partir de descriptions textuelles, ou pour convertir de l'audio en texte ou vice-versa.

Les bienfaits

Comme l'ordinateur personnel, l'IA a de nombreuses utilisations pratiques. Elle peut alléger la charge de travail, réduire le temps nécessaire à l'accomplissement de tâches et améliorer la productivité. Par exemple, l'IA/AAA peut réduire radicalement le temps passé par les médecins à rédiger des comptes rendus de visites de patients et peut également les résumer en traduisant la terminologie médicale dans un langage compréhensible par les patients. Elle peut également analyser les radiographies pour faciliter les diagnostics médicaux.

Les instituteurs ont utilisé l'IA comme assistant pédagogique – bien que certains signalent que cette application doit être améliorée – et les écrivains l'ont utilisée pour établir des fiches de lecture et des avant-projets. Les agents immobiliers l'ont trouvée indispensable pour rédiger des annonces, calculer les paiements hypothécaires et rechercher les utilisations autorisées d'un terrain. Certaines villes réduisent les embouteillages en utilisant l'IA pour ajuster les feux de circulation en temps réel.

Pour faire face à la crise climatique, les algorithmes d'IA sont déjà utilisés pour prévenir les fuites de méthane des gazoducs, améliorer les technologies de stockage des batteries, rendre le fret et le transport plus efficaces, planter des forêts à l'aide de drones et réduire la consommation d'énergie dans les bâtiments, parmi de nombreuses autres tâches. A terme, ils pourraient être utilisés pour accélérer le développement de la fusion nucléaire, pour détecter les changements dans le pergélisol et même pour extraire directement les gaz à effet de serre de l'atmosphère.

Les effets néfastes

Certaines applications de l'IA présentent toutefois des inconvénients, en particulier celles qui impliquent des processus de sélection humaine. Bon nombre de ces problèmes sont dus aux données biaisées utilisées pour entraîner le modèle d'une IA. Dans le système médical, par exemple, les algorithmes d'IA jouent un rôle dans les décisions concernant la distribution des organes et d'autres éléments des soins de santé, et ces biais ont parfois

un pouvoir de vie ou de mort.

Un problème connexe est la tendance du public à accepter les conclusions découlant de l'IA comme des certitudes plutôt que comme des probabilités. Cela est particulièrement problématique dans la mesure où les décisions judiciaires en matière de condamnation sont de plus en plus souvent prises par des algorithmes fermés qui tentent d'évaluer le risque qu'un prévenu commette d'autres crimes à l'avenir. De plus, il n'existe actuellement aucune mesure permettant de lutter contre les préjugés raciaux dans le domaine de l'IA. En l'absence de lignes directrices en matière de transparence et de responsabilité quant aux données ou aux algorithmes d'IA qui les utilisent, les citoyens qui se sentent jugés de manière erronée par des systèmes d'IA ne disposeraient pas des informations nécessaires pour introduire une action en justice et perdraient donc l'accès à une procédure régulière.

Le danger que l'IA sape la démocratie est une autre préoccupation majeure, et les professeurs de Harvard, Archon Fung et Lawrence Lessig, ont présenté un scénario crédible dans lequel des algorithmes en compétition, plutôt que des candidats, pourraient tout à fait décider de l'issue des élections. Ainsi, les modèles d'IA générative comme ChatGPT seraient efficaces pour manipuler le comportement par micro-ciblage car : 1) ces modèles peuvent générer d'innombrables messages uniques pour une personne au cours d'une campagne ; 2) l'apprentissage par renforcement peut générer une série de messages qui deviennent de plus en plus susceptibles de changer le vote d'une personne ; et 3) l'algorithme pourrait mener des « conversations » dynamiques avec des millions d'individus au fil du temps. Le seul objectif de la machine étant de maximiser la part des votes, elle pourrait concevoir des stratégies uniques pour y parvenir, où ni la vérité ni l'exactitude n'entreraient en ligne de compte. En outre, les personnes recevant ces messages n'auraient aucun moyen de savoir qu'elles auraient été dupées.

De plus, comme l'IA génère des faux – appelées par euphémisme « hallucinations » – en réponse à des requêtes, cela exclut son utilisation dans des applications critiques telles que la recherche juridique. Un problème connexe est sa capacité à créer des *« deepfakes »* des faux numériques vraisemblables, pouvant montrer ou faire dire n'importe quoi et qui peuvent avoir des effets dévastateurs sur la vie de personnes ciblées. La désinformation est facile à créer et à faire circuler (certains disent que c'est comme des médias sociaux sous stéroïdes), et des applications telles que la reconnaissance faciale se sont révélées suffisamment imprécises pour que les arrestations injustifiées deviennent monnaie courante – et, là encore, difficiles à réfuter.

A l'échelle mondiale, les algorithmes d'IA pourraient être utilisés par des terroristes pour créer des logiciels malveillants sophistiqués ou des armes biologiques, ce qui aurait des conséquences désastreuses pour la société.

L'homme et la machine

> « Des machines de plus en plus sophistiquées assureront la production. Le problème du chômage cédera la place à celui des loisirs, dont l'organisation sera d'importance capitale. Finalement, toute la production de notre civilisation sera assurée par des machines, libérant l'être humain pour lui permettre d'explorer sa véritable nature et sa raison d'être. Avec le temps, ces machines seront produites par un acte de volonté créatrice de l'être humain. Jusqu'à présent, nous n'avons fait qu'effleurer le potentiel du mental humain. »
>
> [B. Creme, *La Mission de Maitreya*, tome 1]

Le pire serait-il à venir ?

La prochaine étape de l'IA actuellement en débat est l'intelligence artificielle générale (IAG). Elle n'existe pas encore, mais GPT-4 d'OpenAI (soutenu par Microsoft) est décrit comme ayant des *« étincelles d'intelligence générale avancée »* ; il s'agit donc d'un précurseur. C'est ici que la controverse sur le développement de l'IA devient sérieuse.

Une fois que l'IA pourra s'améliorer et devenir super-intelligente (c'est-à-dire capable de surpasser les humains), nous nous trouverons en territoire inconnu. Certains craignent que des robots super-intelligents soient capables de se reproduire et de s'améliorer à un rythme surhumain, de neutraliser facilement les protections mises en place par l'homme et de dominer ou de détruire l'humanité. Il s'agit peut-être d'un rêve éveillé, puisque l'IA n'est

pas consciente ; cependant, entre de mauvaises mains, l'IAG pourrait certainement être manipulée pour causer des ravages dans de vastes secteurs de la civilisation humaine.

L'un des inconvénients bien connu de cette approche – outre les inconnues susmentionnées – est la quantité d'énergie nécessaire à l'entraînement des modèles. Les chercheurs ont constaté que la formation d'un grand modèle libère environ 284 tonnes de CO2, contre 5 tonnes d'équivalents CO2 par an par personne en moyenne. L'utilisation généralisée de ces modèles gourmands en énergie ne serait pas bénéfique pour l'environnement.

La raison pour laquelle des entreprises comme OpenAI et Google DeepMind veulent créer l'IAG est également troublante, au-delà de l'appât du gain. Comme l'indique un article récent d'Emile P.Torres sur le site web Truthdig, la vision du monde qui sous-tend la course à la création de l'IAG est une vision « techno-utopique » de l'avenir, qui inclurait la transformation de l'espèce humaine pour créer une nouvelle race supérieure de « post-humains » (pensez à l'eugénisme) et potentiellement de milliards d'« humains numériques » désincarnés.

Le cœur du dilemme – pour les formes actuelles mais surtout futures d'IA – est en fait l'humain. C'est-à-dire le niveau moyen de conscience humaine. De toute évidence, tout outil peut être utilisé à bon ou mauvais escient, selon l'intention de son utilisateur : un couteau peut servir à tuer ou à couper des légumes. Les médias sociaux peuvent être utilisés pour recruter des militants pour lutter contre l'injustice, ou pour fomenter la haine et la violence. Il en va de même pour l'IA.

Ne pas ouvrir davantage la boîte de Pandore

En définitive, l'humanité ne devrait pas créer de technologies puissantes et potentiellement dangereuses tant qu'elle n'a pas la conscience, l'intégrité morale et le développement évolutif nécessaires pour prendre des décisions judicieuses quant à leur utilisation. Un simple coup d'œil à notre « civilisation » actuelle montre que nous ne sommes pas assez évolués pour être inoffensifs les uns envers les autres ou envers les autres règnes de la nature. Par conséquent, on ne peut pas faire confiance à l'IAG. Si la technologie est disponible, elle sera utilisée – et certainement par au moins quelques mauvais esprits. Nous avons transformé des outils puissants en armes par le passé et nous continuerons à le faire à l'avenir tant que nous n'aurons pas pris conscience de ce que nous sommes vraiment.

L'apparition publique de Maitreya, l'Instructeur mondial pour cette ère, prédite par l'ésotériste Benjamin Creme dans son livre *La Réapparition du Christ et des Maîtres de Sagesse,* accélérera notre développement évolutif et nous enseignera que nous faisons tous partie d'une grande famille humaine – des dieux en incarnation. Maitreya expliquera clairement que la création d'une existence utopique ne nécessite pas d'IA, mais seulement la découverte et le développement de nos capacités inhérentes en tant qu'êtres divins et l'application des principes spirituels enseignés par les sages tout au long de l'histoire.

L'IA n'existait pas encore lorsque B. Creme a écrit son livre, mais évoquant d'autres expériences scientifiques dans lesquelles des savants manipulaient le mécanisme des forces évolutives (ingénierie biogénétique), B. Creme affirme que Maitreya a lancé un avertissement selon lequel il est dangereux pour les scientifiques d'utiliser leurs connaissances pour *« jouer à Dieu »* : *« Vous êtes capable de réaliser certaines choses,* explique-t-il*, mais vous ne devriez pas les faire. »* Le même avertissement pourrait et devrait certainement s'appliquer à l'IAG.

La voie la plus prudente et la plus juste consisterait à réglementer toutes les formes d'IA actuellement utilisées afin de : 1) garantir que les données biaisées, les « hallucinations » et la confiance excessive dans les décisions de l'IA ne portent préjudice à personne, en particulier aux plus vulnérables ; 2) interdire certaines utilisations de cette technologie, telles que les campagnes politiques ou l'utilisation des armements ; et 3) imposer des systèmes de sécurité solides là où les modèles d'IA sont développés et utilisés, afin d'éviter qu'ils ne tombent entre les mains de personnes mal intentionnées. Enfin, il serait sage d'arrêter tout développement de l'IA pour une durée indéterminée, de peur d'ouvrir davantage la boîte de Pandore.

Auteur : Cher Gilmore, collaboratrice de Share International basée à Los Angeles, en Californie (États-Unis)
Sources : Stanford's One Hundred Year Study on Artificial Intelligence (Une année d'études sur l'intelligence artificielle), New York Times, The Future We Choose, par Christiana Figueres et Tom Rivett-Carnac.
Thématiques : Sciences et santé
Rubrique : De nos correspondants