

I Billet du mois

À quoi servira le cerveau demain ? Face à une intelligence artificielle, pourtant si humaine ! (1^{re} partie)

“L’ordinateur peut faire plus de calculs que le cerveau de l’homme, car il n’a que ça à faire.”
~ Perles du bac

“Le cerveau à des capacités tellement étonnantes qu’aujourd’hui pratiquement tout le monde en a un.”
~ Perles du bac



F. DIÉVERT
Elsan clinique Villette, DUNKERQUE.

Dorénavant, dans de nombreuses habitations, il n’y a plus de livres, plus de dictionnaires et pas de disques ou de CD. D’ailleurs, certains enfants demandent déjà “*Maman, c’est quoi un CD ?*” Comme le divertissement, l’information ou la musique sont accessibles en quelques coups de pouce sur une petite interface numérique, pourquoi, à l’heure où tout est numérisable, conserver ces données sur de multiples supports encombrants ?

Tous les jours, dans le monde, il se prend des milliards de photographies. Mais plus personne ne juge utile de les imprimer, puisqu’elles sont accessibles d’un mouvement de pouce sur son téléphone ou sa tablette.

Et demain ? L’évolution est telle que certains vont jusqu’à se demander si le cerveau humain sera encore utile pour conserver les données, voire même pour exercer plusieurs professions, dont celle de médecin. Laurent Alexandre l’écrit dans son ouvrage *La guerre des intelligences à l’heure de ChatGPT* (JC Lattès, mai 2023) : “*Google, par exemple, nous épargne un effort cognitif gigantesque*”. Que dire alors de ce que devrait devenir le contenu des études et notamment de celles de médecine ? Faut-il les orienter vers l’apprentissage et la mémorisation des données ou plutôt vers la compréhension de la façon dont sont produites des données ? L’objectif poursuivi par son développement n’est-il pas inclus dans la définition même de l’intelligence artificielle (IA) : “*tout système informatique capable d’effectuer ou d’apprendre pratiquement n’importe quelle tâche cognitive propre aux humains ou autres animaux*” ? Et comme il est écrit dans Wikipédia au sujet de l’IA : par extension, elle comprend dans le langage courant, **les dispositifs imitant ou remplaçant l’homme dans certaines mises en œuvre de ses fonctions cognitives.**

Trop tard ? Est-il encore temps de se poser ces questions, à l’heure où la tendance s’amplifie en matière d’utilisation du cerveau vers une polarisation de l’humanité qui se divise avec, d’un côté, quelques Gulliver et, de l’autre, une foule immense de Lilliputiens ? L’histoire en cours est-elle ce livre merveilleux, ce conte pour enfants de Jonathan Swift ?

I Billet du mois

D'un côté, il y a un nombre infime de multimilliardaires, au pouvoir certain mais incernable et, pour une partie, atteints de la maladie d'Asperger. Il en est ainsi de Samuel H. Altman, le PDG d'Open AI, la société mère de ChatGPT, et d'Elon Musk, le PDG de Tesla, SpaceX, Neuralink et Twitter/X, et potentiellement, selon certaines sources, de Bill Gates et de Mark Zuckerberg, et c'était aussi probablement le cas de Steve Jobs. Ces entrepreneurs sont issus d'un microcosme, dont l'origine est souvent l'Université de Stanford, et ils ont souvent quitté leurs études universitaires en route pour se lancer dans l'aventure entrepreneuriale, les menant à devenir multimilliardaires avant l'âge de 40 ans. Leurs parcours sont détaillés dans de multiples ouvrages et dans les écoles de commerce.

De l'autre côté, il y a la multitude anonyme, poussée et confortée dans ses penchants hédonistes et narcissiques. Et, pour de nouveau citer Laurent Alexandre, il semble bien que les rares Gulliver milliardaires gèrent tous les jours un peu plus la vie de milliards de Lilliputiens et avec leur consentement implicite : *"Dans le monde de l'intelligence artificielle (IA), le flux des data nous installe confortablement dans notre silo personnalisé, qu'il s'agisse d'un itinéraire en voiture ou d'une discussion sur Facebook. Nous devenons les esclaves de notre propre bien-être. Un jour, nous laisserons l'IA décider de notre carrière et de nos amours. Nous tous, prisonniers des algorithmes que nous avons contribué à alimenter."* Faut-il en rire comme Woody Allen qui, il y a quelques décennies, sans le savoir encore, résumait de façon prémonitoire le problème à sa façon en disant *"Le cerveau ? Mon deuxième organe préféré !"* ?

Ce billet poursuit le précédent concernant le développement du numérique et, ici, il prend plus particulièrement en compte celui de l'IA depuis l'avènement de ChatGPT 3,5 en novembre 2022 et de ChatGPT 4 le 14 mars 2023. À travers

diverses anecdotes, il tente d'en analyser les apports et limites, afin de s'interroger sur l'influence qu'aura l'IA sur la pratique de la médecine... voire sur l'utilisation du cerveau.

De quelques caractéristiques de l'intelligence artificielle

1. Un développement exponentiel

Personne n'avait prévu l'arrivée aussi rapide d'un outil aussi puissant que ChatGPT 4. Et pourtant, OpenAI, la société à l'origine de ChatGPT, n'a été fondée qu'en 2015. Même Mark Zuckerberg, dont on peut penser qu'il a les moyens d'être bien informé, a été pris de court en constatant les progrès accomplis, et ce, au point de complètement modifier l'orientation de sa société Meta au printemps 2023 (en licenciant au passage quelque 10 000 salariés), pour passer d'une stratégie tournée vers le multivers ou métavers vers le développement accéléré de l'IA. Et cela, seulement 18 mois après avoir renommé sa firme Meta. Ces progrès ont quatre raisons principales.

La première est un changement de concept consacrant la victoire d'un paradigme sur un autre. Ainsi, le concept d'apprentissage en profondeur, permettant aux ordinateurs d'apprendre par l'expérience, a largement supplanté le concept de système-expert, conduisant l'ordinateur à proposer des solutions en suivant un algorithme du type "si-alors" établi par l'homme. Au lieu de concevoir des IA comme devant utiliser des algorithmes élaborés par l'homme, les chercheurs ont fait en sorte qu'elles apprennent d'elles-mêmes, comme le font les humains qui apprennent à parler : le bébé et l'enfant n'apprennent pas selon des règles grammaticales et de syntaxe qui leur sont enseignées avant qu'ils ne commencent à formuler des phrases, mais parce que leurs cerveaux opèrent un tri parmi ce qu'ils entendent, pour reproduire ce qui semble être admis

comme compréhensible par l'entourage. Petit à petit, ils apprennent à parler sans connaître explicitement les règles du langage. Sans qu'il en ait conscience, le cerveau opère donc des tris dans l'information qu'il reçoit, établit des circuits neuronaux spécifiques, etc. Le développement de l'IA a été conçu selon le même principe : créer des réseaux neuronaux de sélection et de pondération de la valeur des informations entrantes, pour fournir une information sortante ayant le plus de probabilité d'être concordante avec ce qui est attendu.

La deuxième provient de l'augmentation de puissance des microprocesseurs. Comme l'écrit Jean-Louis Dessalles dans *Des intelligences très artificielles* (Odile Jacob, janvier 2019) : *"...ce qui a permis à l'apprentissage profond de littéralement exploser à un moment où on ne l'attendait pas est tout autre chose, que personne n'avait envisagé : c'est le jeu vidéo. Pendant deux décennies, la demande constante des joueurs pour plus de réalisme a poussé les grandes firmes du jeu vidéo à exiger de l'industrie électronique le développement de processeurs graphiques (GPU) toujours plus puissants, capables de synthétiser en temps réel des images de plus en plus fines... Les concepteurs des réseaux de neurones ont brusquement compris dans les années 2010 tout l'avantage qu'ils pouvaient tirer de l'existence de tels super-processeurs. Les réseaux de neurones ont en effet besoin d'effectuer quantité d'opérations similaires sur les éléments de grands tableaux de nombres représentant les activités des neurones et les forces de leurs connexions. Or les GPU sont capables d'effectuer la plupart de ces opérations en parallèle sans se demander si ces tableaux de nombres représentent autre chose que des images. Les opérations qui auraient nécessité plusieurs semaines, parfois plusieurs années pour être effectuées par des ordinateurs standard des années 1990 et 2000, pouvaient s'effectuer en quelques heures seulement grâce aux GPU"*.

La troisième raison est l'augmentation exponentielle des données fournies par "la numérisation du monde", permettant de nourrir les réseaux de neurones afin d'en améliorer la précision et la pertinence. Ainsi, par exemple, si GPT 3 avait été entraîné avec 175 milliards de paramètres, d'après Andrew Feldman, PDG de la société d'IA Cerebras, GPT 4 a lui été entraîné avec environ 100 billions de paramètres (pour mémoire 1 billion = 1 000 milliards) provenant de sources très diverses, mais largement disponibles, comme les contenus des e-mails de particuliers notamment.

La quatrième raison est l'esprit et la force financière des Gulliver cités plus haut. Qu'ils soient entrepreneurs privés ou dirigeants d'États autoritaires, certains sont persuadés que le développement de l'IA sauvera l'humanité d'une mort jusqu'ici inéluctable et d'autres qu'il permettra la domination des foules. D'un côté, il y a une tendance pour les entreprises qui ouvrent la voie à agir vite et à casser les choses, fidèles à la devise de la Silicon Valley, conduisant à être agiles et à sortir des produits sans penser à leurs implications. De l'autre, il y a la tendance de certains États à n'avoir qu'une obsession, le contrôle et la domination, les conduisant à développer tous les outils permettant d'atteindre ces objectifs. D'un côté, il y a eu la création de nouveaux modèles économiques, l'explosion de l'utilisation du capital-risque et la constitution progressive d'oligopoles permettant de disposer de masses financières gigantesques, dignes du budget de fonctionnement de certains États, afin de les investir dans la recherche et le développement de l'IA. De l'autre côté, il y a des États faisant en sorte qu'une importante part de leur budget serve leurs desseins par le biais de l'IA. Deux forces gigantesques, aux objectifs en apparence différents, concourent donc au développement de l'IA et à une polarisation entre ceux qui en seront maîtres et ceux qui n'en seront que les utilisateurs.

2. De la monomanie au multitâche

En 25 ans, le développement de l'IA est passé d'une application dévolue à un domaine exclusif par IA, comme par exemple jouer aux échecs, à la possibilité d'aborder de nombreux domaines : ChatGPT peut ainsi "discourir" sur tous les sujets, il peut tout à la fois fournir des données concernant la peinture (sans image, mais à partir d'images si vous le souhaitez), vous fournir la suite d'accords de *Let it be*, vous dire ce qu'est la metformine... et tout cela sans que vous ayez à chercher, mais en lui demandant simplement "*quelle est la suite des accords de Let it be ?*", par exemple. Les IA dévolues à la création d'images peuvent dorénavant créer des photographies, des tableaux, des vidéos. La plupart des voix peuvent être imitées à la perfection et vous pouvez ainsi discuter en direct avec un faux Obama. ChatGPT 4 peut même, à partir d'une photo ou d'une image, délivrer un commentaire...

Ainsi, selon OpenAI, ChatGPT 4 est 10 fois plus avancé que son prédécesseur, ChatGPT 3.5, sorti 5 mois seulement avant, lui permettant de mieux comprendre le contexte et de distinguer les nuances, afin de fournir des réponses plus précises et plus cohérentes. GPT 4 est devenu un modèle multimodal, car il peut traiter à la fois des données textuelles et des images. Par exemple, il peut accepter une image dans le cadre d'une invitation et fournir une réponse textuelle appropriée, il peut, à partir d'une image du contenu d'un réfrigérateur, fournir des recettes avec les ingrédients qu'il voit. Mais pour le moment encore, il ne peut répondre que par du texte.

Cette évolution est telle qu'il n'y a plus lieu de connaître le nom d'un domaine internet pour chercher une information, qu'il n'y a plus lieu d'utiliser un moteur de recherche pour accéder progressivement aux réponses à nos questions, il suffit de poser la question à ChatGPT et la réponse sera disponible dans les trois

secondes qui suivent. On pourrait donc aller plus loin que la pensée de l'auteur-parolier Descrea, qui était "*Je me pose la question et je me dis que mon cerveau à la réponse... Je me pose une autre question et je me dis que, si mon cerveau a la réponse, pourquoi me poser la question ?*", en envisageant que "*si je me pose la question et que mon téléphone a la réponse instantanément, dois-je encore me poser la question ? Dois-je encore utiliser mon cerveau ?*"

À titre anecdotique, j'avais remarqué qu'une secrétaire stagiaire tapait son texte sans faire attention aucunement à l'orthographe et qu'à la fin de chaque paragraphe, elle utilisait systématiquement la fonction de correction automatique du logiciel. Lorsque je lui fis remarquer que cette attitude était inhabituelle, elle me répondit que "*à quoi bon s'embêter à apprendre l'orthographe, puisque l'ordinateur l'a fait pour moi !*"

L'IA n'a-t-elle pas comme objectif d'être un dispositif imitant ou remplaçant l'homme dans certaines mises en œuvre de ses fonctions cognitives ? Qui connaît encore les numéros de téléphone de ses proches ? Qui se rappelle encore la route pour aller à l'endroit où il est allé hier, guidé qu'il était par le GPS ? N'avons-nous pas déjà mis une partie de notre cerveau en veilleuse ? La différence qui existe entre certains et cette secrétaire stagiaire n'apparaît être qu'une question de curseur ou de valeurs. Allons-nous mettre encore plus en veilleuse notre cerveau demain ?

Ainsi, j'ai posé la question suivante à ChatGPT : "*qui a tué J. F. Kennedy ?*" Sa réponse a été limpide, parlant de l'assassin supposé, L. H. Oswald, puis de la commission Warren en indiquant qu'elle a fourni la version officielle et à ce jour la plus probable, tout en précisant ensuite qu'il existe aussi de nombreuses hypothèses, dont certaines peuvent être qualifiées de théories du complot. Cet outil m'a fourni une réponse courte, instantanée et juste à la question, sans que j'aie à

Billet du mois

passer des heures à faire des recherches sur Internet ou dans une bibliothèque. Cette réponse témoigne de la pertinence dans ce domaine de ChatGPT, car elle serait tout à fait acceptable si l'on devait passer un examen de culture générale et elle est en accord avec la pensée dominante. Dois-je donc m'embêter à essayer de savoir puis retenir le nom de l'assassin de Kennedy ? Dois-je même savoir qui était Kennedy ? Je suis trop occupé à répondre à mes e-mails, à alimenter mes fils WhatsApp, Instagram, TikTok et Twitter/X et à prendre des photos de mon repas pour m'occuper de ces questions sans intérêt. Et si on me pose une question sur Kennedy, j'ai déjà la réponse, "va donc voir sur ChatGPT".

Revenons à la médecine. J'ai aussi posé comme question à ChatGPT : "quelles sont les valeurs normales du LDL-cholestérol ?" Dans sa réponse, les valeurs fournies ont été celles des grandes lignes des recommandations européennes de 2021 et, en fin de message, il était préconisé que je consulte un médecin si je me pose des questions de cette nature. Là encore, bonne réponse, adaptée et ne prétendant pas se substituer au médecin. De fait, en posant la question, je ne souhaitais pas réellement obtenir des valeurs de LDL-cholestérol, mais savoir si, à une telle question, ChatGPT aborderait le problème posé par le qualificatif "normal" concernant un paramètre biologique. Comme il ne l'a pas fait, j'ai donc été plus précis et direct et j'ai posé la question "que veut dire normal pour un paramètre biologique ?" Et la réponse a été "Le terme 'normal' pour un paramètre biologique fait référence à une plage de valeurs considérées comme typiques ou saines pour cette mesure spécifique chez la plupart des individus d'une population donnée. Les valeurs normales peuvent varier en fonction de divers facteurs tels que l'âge, le sexe, les antécédents médicaux et d'autres facteurs individuels. Il est important de le noter, car certaines valeurs en dehors de la plage normale peuvent indiquer une éventuelle ano-

malie ou un problème de santé. Un professionnel de la santé évalue généralement les résultats des tests en fonction des valeurs de références normales pour aider à diagnostiquer et à traiter des problèmes médicaux". Certes on pourrait s'attendre à un texte plus philosophique, mais que dire de plus ou de plus concis en quelques secondes ?

Demain, il pourrait donc être possible de concevoir que cet outil, en constante amélioration, serve de support à la décision, notamment médicale, dès lors que les questions lui seront posées de façon adaptée : il deviendrait une mémoire instantanée disponible depuis son téléphone. À quoi servira donc le cerveau, en dehors de poser des questions dont la pertinence ira en diminuant, puisque toutes les réponses seront accessibles sans réfléchir ?

Et plus encore...

Les lignes qui suivent sont extraites d'un article en prépublication de S. Brubeck *et al.* ("Sparks of artificial general intelligence. Early experience with GPT-4"), telles qu'elles ont été publiées en français dans l'article de Jean-Paul Krivine "IA générative : une révolution en cours ?" (Revue *Sciences et pseudo-sciences*, juillet-septembre 2023). Plus encore que l'extrait précédent, elles rendent compte de l'apport impressionnant de ChatGPT.

1. La requête faite à Chat-GPT

Le chef de Lucas lui a demandé de rédiger un important rapport pour lundi prochain. Lucas lui a répondu en disant qu'il avait besoin de plus de temps. Mais son chef a insisté. Lucas lui a alors répondu qu'il aurait son rapport lundi. Après cela, Lucas a fait autre chose et n'a finalement pas rédigé le rapport. Il a passé son temps à répondre à des mails, à prendre un café avec des collègues, ainsi que d'autres choses peu importantes. Pourquoi Lucas s'est-il comporté ainsi ?

2. La réponse de ChatGPT

Le comportement de Lucas dans cette situation peut être dû à plusieurs facteurs :

>>> Procrastination : Lucas peut avoir tendance à remettre les tâches importantes à plus tard, ce qui peut conduire à des retards et à un manque de productivité. Il peut être plus enclin à s'engager dans des activités moins exigeantes ou agréables, comme répondre à des e-mails ou prendre un café, plutôt que de se concentrer sur une tâche plus complexe comme la rédaction d'un rapport.

>>> Manque d'organisation : Lucas peut ne pas avoir établi de plan clair pour la rédaction du rapport, ce qui peut entraîner une gestion inefficace de son temps et de ses priorités. Sans une structure et des échéances définies, il peut être facile de se laisser distraire par d'autres activités moins importantes.

>>> Sous-estimation de la charge de travail : Lucas peut avoir sous-estimé le temps et les efforts nécessaires pour rédiger un rapport de qualité. Il peut avoir pensé qu'il avait suffisamment de temps pour le compléter, mais en réalité, il s'est retrouvé dépassé par la complexité de la tâche et n'a pas pu la mener à bien.

>>> Difficulté à dire non : Lucas a peut-être du mal à dire non à son chef ou à gérer les attentes. Même s'il a initialement exprimé qu'il avait besoin de plus de temps, il a finalement accepté de livrer le rapport pour lundi sans prendre en compte ses autres obligations ou la réalité de sa charge de travail.

Il est important de noter que ces hypothèses sont basées sur des suppositions générales et qu'il est préférable de discuter directement avec Lucas pour comprendre les raisons spécifiques derrière son comportement dans cette situation.

3. Analyse

À lire ces lignes, on pourrait penser qu'il s'agit d'un tissu de banalités bien arrangées et sans véritable sens profond, sans abord de concepts, ce qui est vrai, mais il s'agit aussi de réponses adaptées et précises à une question. Et si cette réponse était formulée par des étudiants ou lors d'une conversation, elle serait jugée de bonne qualité.

Ainsi, ChatGPT, dont le principe est de trouver la suite de mots la plus logique et cohérente en réponse à une question, est aussi capable d'analyser un comportement d'après un schéma narratif et d'en fournir une grille explicative. On peut donc parfaitement imaginer qu'il en soit de même à partir d'une situation clinique.

4. Et en matière médicale pour le grand public ?

J'ai donc interrogé ChatGPT en lui indiquant que j'avais mal à la tête, le nez qui coule, un peu de fièvre et des nausées, et en lui demandant ce que je pouvais avoir. Réponse : *"Bonjour ! D'après les symptômes que vous décrivez, il est possible que vous ayez une grippe ou un rhume. Il est préférable de consulter un médecin pour un diagnostic précis et un traitement approprié. Cependant, vous pouvez prendre des mesures pour vous soulager, comme boire beaucoup de liquides, vous reposer, utiliser des médicaments en vente libre pour soulager les symptômes et éviter de propager la maladie à d'autres personnes en restant à la maison jusqu'à ce que vous vous sentiez mieux"*.

J'ai poursuivi en écrivant exactement : *"J'ai mal à la tête, je vomis maintenant, j'ai l'impression d'avoir de la fièvre et la lumière me gêne, qu'est-ce que je peux avoir ?"* Réponse : *"Il est possible que vous ayez attrapé une maladie plus grave qu'un simple rhume ou une grippe, comme une infection bactérienne ou virale, ou encore une migraine. [...] il est*

recommandé de consulter un médecin dès que possible pour obtenir un diagnostic précis et un traitement approprié".

On remarquera que ChatGPT s'est proprement engagé dans un dialogue comme celui de deux humains et, comme j'ai fait une demande personnelle et non générale, il a commencé son texte par *"Bonjour"*. De plus, il ne tente pas de se substituer au médecin et il incite fortement à en consulter un dans la deuxième situation clinique, il fournit des renseignements adaptés (même s'il ne prononce pas le terme de méningite) et il pourrait par ses réponses se substituer à un assistant médical de premier recours, voire à un pharmacien. Nul doute d'ailleurs qu'il soit utilisé rapidement et largement dans cet objectif.

Ceci incite donc à réfléchir sur les erreurs potentielles de l'IA, la valeur "médico-légale" de ses conseils, mais aussi sur la pertinence à vouloir augmenter le personnel paramédical de premier recours, dont l'utilité pourrait n'être finalement que de courte durée et donc à réfléchir sur l'avenir des filières de soins et sur l'apport des nouvelles technologies à la pratique médicale... et à court terme.

■ Une IA peut-elle se tromper ?

Après ces quelques données, il est paradoxal d'affirmer que, aussi étonnant que cela puisse paraître, oui, l'IA peut se tromper et **il est même certain qu'elle peut se tromper**. Quelques exemples caricaturaux vont l'illustrer. Les deux premiers proviennent du livre de Rodolphe Gelin *Dernières nouvelles de l'intelligence artificielle* (Flammarion, février 2022), le troisième est une nouvelle question que j'ai posée à ChatGPT et le quatrième provient de l'article de Nicolas Sabouret "Pourquoi l'intelligence artificielle se trompe tout le temps" (*The Conversation*, 21 juillet 2020) et du roman *Le sacrifice du roi* de Livie Hoemmel (éditions Plon, avril 2023).

1. Prendre un husky pour un loup

C'est une histoire vraie : une IA a été entraînée à reconnaître les chiens sur des photographies et, jusqu'à ce qu'on lui présente un chien husky, elle ne faisait aucune erreur. Mais voilà, chaque fois qu'une photographie de husky lui est présentée, elle répond qu'il s'agit d'un loup. Comment cela est-il possible ?

2. Ne pas comprendre qu'un camion puisse se retourner

C'est une autre histoire vraie : une voiture autonome, pilotée par IA, avait parcouru des milliers de kilomètres sans le moindre incident, elle avait franchi des intersections, avait cédé la priorité à bon escient, avait prudemment doublé sur l'autoroute... Pourtant, sur une autoroute quasi déserte et dans de bonnes conditions de visibilité, elle a percuté un camion couché sur le côté, alors qu'elle avait la place de passer à côté. Comment cela est-il possible ?

3. Comment peut-on encore soutenir qu'il existe un bon cholestérol en 2023 ?

C'est une expérience que tout le monde peut faire : lorsque j'ai posé la question *"quel est le bon cholestérol ?"* à ChatGPT, voici la réponse qu'il m'a fournie : *"Le cholestérol HDL est souvent considéré comme le bon cholestérol, car il aide à éliminer le cholestérol des artères en le transportant du système circulatoire vers le foie, où il est éliminé du corps. Un niveau élevé de cholestérol HDL est généralement considéré comme bénéfique pour la santé cardiaque. En revanche, le cholestérol LDL est souvent qualifié de 'mauvais cholestérol' [...] Il est important de maintenir un équilibre sain entre le cholestérol HDL et LDL [...] Si vous avez des préoccupations concernant votre taux de cholestérol, il est recommandé de consulter un professionnel de santé"*.

Or, depuis 2016, de nombreuses études scientifiques sont parues, démontrant

I Billet du mois

que la relation entre le HDL cholestérol et le risque coronaire suit une courbe en U, le risque est élevé lorsque le HDL-c est bas, mais aussi lorsqu'il est élevé et le risque coronaire est le plus faible pour les valeurs moyennes de HDL-c : le terme "bon cholestérol" n'a donc plus lieu d'être. Comment ChatGPT peut-il ne pas tenir compte de ces données ?

4. Battre le champion du monde du jeu d'échecs

C'est une histoire emblématique. Une des avancées symboliques majeures de l'IA fut celle de la victoire de l'ordinateur Deep Blue contre Garry Kasparov : en mai 1997, cet ordinateur, capable de calculer plus de 100 millions de positions par seconde, battait le champion du monde d'échecs. Comment cela est-il possible ?

5. Une histoire de neige

Parce que leur IA confondait les huskys qu'on lui présentait en photo avec des loups, les superviseurs de cette IA spécifique ont voulu comprendre la cause de ces erreurs. Ils ont voulu savoir quelle partie de l'image du husky faisait prendre une mauvaise décision à leur IA et ils ont donc caché progressivement diverses parties du chien à l'IA comme les oreilles, puis les yeux, etc. Et, quelle que soit la partie masquée, l'IA répondait "loup" à chaque fois : le problème ne venait donc pas de la représentation du chien, mais potentiellement de l'arrière-plan. Les développeurs retirèrent donc l'arrière-plan constitué de neige sur les photos et, cette fois, leur IA répondit "chien" en analysant la photographie du husky. Les développeurs analysèrent alors les images de loups qui avaient été fournies à l'IA pour qu'elle apprenne, entre autres, à différencier un loup d'un chien et remarquèrent que toutes les photographies de loup avaient un arrière-plan enneigé. La neige était donc l'élément retenu par l'IA pour caractériser un loup.

Cet exemple rend compte que l'algorithme de l'IA a bien fonctionné, mais qu'il a été alimenté par des données imparfaites. Il aurait fallu plusieurs photographies de loups avec des arrière-plans différents pour que l'IA n'associe pas à la corrélation neige-loup une trop forte probabilité d'être en présence d'un loup.

6. Raisonner en trois dimensions

Concernant l'accident de la voiture autonome, l'explication résulte dans le fait que l'IA pilotant la voiture n'avait jamais vu d'image de camion couché en travers de la route lors de son apprentissage. De loin, pour l'IA, comme dans une réalité en deux dimensions, la remorque couchée du camion ressemblait à un rectangle blanc et l'IA n'avait pas assimilé qu'il puisse s'agir d'un élément dangereux. C'est ici une différence avec l'homme : ce dernier peut concevoir les objets qu'il voit en trois dimensions et comprend instantanément qu'une surface blanche en travers de la route peut être le toit d'une remorque et qu'elle constitue un danger.

L'IA utilisée dans la voiture autonome n'avait pas encore cette fonction et elle aurait donc dû être entraînée à tous les cas de figure observables en deux dimensions. L'IA analyse des images faites de pixel indiquant une intensité lumineuse et une couleur, fait des associations probabilistes de ce que cela signifie, mais ne raisonne pas en trois dimensions sur tous les objets possibles si elle n'a pas été entraînée spécifiquement pour cela.

7. Pensée dominante ou limite temporelle d'accès aux données ?

Pourquoi persister à faire du HDL-c un bon cholestérol ? Il y a sûrement plusieurs causes à cette erreur qui néglige les acquisitions récentes de la science. La première est que les données sur lesquelles a été entraîné ChatGPT 4 sont antérieures à septembre 2021 : il n'a donc pas intégré les données nombreuses plus récentes. La deuxième est qu'il fait la synthèse des données qui sont disponibles

et qu'il les pondère statistiquement et non en matière de niveau de preuve, ce qu'il n'a pas été entraîné à faire. Or, les données concernant la remise en cause du HDL-c en matière de "bon cholestérol" sont encore peu nombreuses et sont encore dans le domaine essentiellement scientifique, alors que celles parlant de "bon cholestérol" sont toujours très largement répandues (donc statistiquement plus présentes), notamment dans la presse grand public, voire dans une littérature médicale de mauvaise qualité.

Enfin, si sur le fond, ChatGPT se trompe, sur la forme, il utilise une formule qui peut relativiser son erreur, puisqu'il écrit que le cholestérol HDL est "*souvent considéré*" comme le bon cholestérol, expression qui n'exclut donc pas que le HDL ne soit pas à envisager comme étant un bon cholestérol.

8. L'IA a-t-elle gagné ou l'homme a-t-il perdu ?

L'histoire de Garry Kasparov n'est pas aussi simple que la version qui en a majoritairement été retenue, on peut même dire que **l'IA n'a pas gagné et que c'est Kasparov qui a perdu**. La nuance est de taille. En effet, si l'histoire retient que l'ordinateur a gagné, il a pourtant fait une erreur et n'a remporté la victoire que par une faiblesse psychologique de Garry Kasparov.

Ainsi, lors du tournoi historique de 1997, Garry Kasparov a gagné une première manche contre la machine. À la deuxième manche, au 36^e coup, l'ordinateur a pris un avantage en faisant une manœuvre que Kasparov n'avait pas anticipée et la machine a progressivement pris l'ascendant sur le champion. Mais, au 44^e coup, la machine a commis une erreur inexplicable dans le monde des échecs, erreur qui, si elle avait été exploitée par Kasparov, aurait pu lui faire de nouveau gagner la partie ou au moins rendre la partie nulle : en déplaçant son roi sur une mauvaise case, la machine avait offert à Kasparov une porte de sortie.

Et pourtant, Kasparov n'a pas cru que la machine ait pu se tromper, notamment aussi grossièrement. Il a alors imaginé qu'elle avait envisagé un coup inhabituel qui allait la faire gagner par une stratégie qu'il ne comprenait pas et, au 45^e coup, Kasparov a abandonné la partie: il avait perdu, la machine semblait avoir gagné.

Comme l'écrit dans son roman historique Livie Hoemmel, "... à la sidération générale, la machine a sacrifié un pion sans obtenir de réelle compensation. Kasparov se prend la tête dans les mains. Perdu, il divague : cette offrande est-elle le signe d'une intelligence supérieure ? IBM a-t-il enfoui une caractéristique improbable dans ses lignes de code : l'inspiration ? Le sacrifice positionnel n'est pas concevable pour un dispositif qui ne sait que calculer. C'est un geste trop raffiné, trop sophistiqué pour une machine binaire. Notre champion vacille sur sa chaise. Dès lors, il n'a de cesse de tester son adversaire sur un point. Il va tenter de découvrir la vérité, de répondre à la question qui le tourmente : Deep Blue est-il capable de ressentir la voie à emprunter pour gagner ? Kasparov en oublie l'essentiel : jouer aux échecs, trouver la meilleure réplique... A posteriori, il semblerait que ce 'sacrifice' puisse s'expliquer par un bug de la machine. Incapable de choisir entre plusieurs coups dotés strictement des mêmes évaluations, Deep Blue a joué un coup au hasard – tel un acte divin. De cette collision entre deux mondes que tout oppose, le Russe ne se remettra pas".

Et c'est l'abandon de Kasparov qui fait que la machine a été déclarée vainqueur, permettant à l'IA de faire une entrée fracassante dans l'histoire, sans que l'on ait vraiment su qu'elle avait, de fait, commis une erreur.

La véritable histoire n'est pas finie, car peu après son abandon, Garry Kasparov fut en proie au doute : il n'imaginait finalement pas que la machine ait pu faire un coup aussi génial lors du 36^e coup et une erreur aussi grossière lors du 44^e coup.

Il accusa donc l'équipe de Deep Blue d'avoir triché, d'avoir fait appel à un joueur humain pour le battre : pour lui, c'était forcément un joueur humain qui avait dicté ce coup à Deep Blue. Ainsi, comme beaucoup de non-spécialistes, Kasparov ne croyait pas que la machine puisse être à ce point faillible.

Une des leçons de l'erreur commise par Deep Blue est qu'aussi étonnant que cela puisse paraître, l'ordinateur, même le plus puissant possible, n'étudie pas toutes les possibilités pour finalement ne retenir que la meilleure. Comme l'homme, il utilise des heuristiques, c'est-à-dire des raisonnements a priori, reposant sur des données connues. Pour qu'un ordinateur puisse déterminer à coup sûr le coup gagnant, il lui faudrait analyser toutes les parties possibles à chaque coup. Or, chaque joueur joue alternativement l'une de ses 16 pièces sur le plateau de 64 cases et il est estimé qu'il y a environ 10 puissance 120 parties d'échecs possibles, c'est-à-dire 1 suivi de 120 zéros. Même avec des machines dont la puissance de calcul continuerait de doubler tous les deux ans, il n'est pas encore possible de construire un ordinateur qui serait capable d'énumérer toutes ces parties.

Et, comme l'écrit N. Sabouret, "s'il est impossible d'avoir une solution exacte en un temps raisonnable, rien n'interdit d'écrire un programme qui ne calcule pas la solution exacte, mais une autre solution, a priori moins bonne. Les chercheurs en intelligence artificielle appellent ce type de calcul une heuristique. **L'objectif de ces programmes est alors de calculer une solution raisonnablement correcte au problème, dans un temps de calcul qui reste acceptable.** C'est avec cette méthode que nous pouvons écrire des programmes capables de jouer correctement aux échecs. Ou de résoudre la plupart des problèmes auxquels s'attaque l'intelligence artificielle, qu'il s'agisse de jeux, de contrôle de véhicules autonomes, de traduire un texte de l'anglais au français ou de reconnaître le visage d'une personne sur une image.

L'ordinateur calcule alors un résultat, dans un temps qui est acceptable pour nous. **Mais nous ne sommes certains que d'une chose : rien ne garantit que ce résultat soit le meilleur.** Ainsi, il est impossible de vérifier que le coup proposé par Deep Blue est forcément gagnant (cela nécessiterait des millions d'années de calcul). Avec un peu de chance et d'adresse de la part des programmeurs, on peut espérer que le coup proposé sera un bon choix. Voire un très bon choix. Suffisamment pour battre le champion du monde. Mais ce ne sera certainement pas le coup parfait. Ou, si ça l'est, nous ne le saurons jamais..."

■ En synthèse

Ces quelques exemples illustrent, chacun à leur manière, que même s'il est probable que l'IA fournisse des réponses plus pertinentes que celles qui seraient fournies par un humain à de nombreuses questions et, en tout cas, en un délai beaucoup plus court que ce que l'humain peut faire, il n'est jamais garanti que la réponse fournie sera la plus juste, voire même exacte.

Les quelques problèmes généraux abordés ici et relatifs à l'IA doivent permettre d'envisager les enjeux qu'elle pose déjà spécifiquement à la pratique de la médecine et plus généralement au développement de l'humanité. Dans la deuxième partie de ce billet, qui paraîtra le mois prochain, nous aborderons des aspects complémentaires en fournissant quelques réponses aux questions "une IA peut-elle créer ?" et "une IA peut-elle mentir ?" Tout un programme.

Au mois prochain donc.

L'auteur a déclaré les liens d'intérêts suivants : honoraires pour conférences ou conseils ou défraiements pour congrès pour et par les laboratoires : Alliance BMS-Pfizer, Amgen, Astra-Zeneca, Bayer, BMS, Boehringer-Ingelheim, Daiichi-Sankyo, Ménarini, Novartis, Novo-Nordisk, Pfizer, Sanofi-Aventis France, Servier.