

R

ENDEZ-VOUS

P.80 Logique & calcul
 P.86 Art & science
 P.88 Idées de physique
 P.92 Chroniques de l'évolution
 P.96 Science & gastronomie
 P.98 À picorer

DERRIÈRE LES MODÈLES MASSIFS DE LANGAGE

Un saut qualitatif a été franchi par l'intelligence artificielle, grâce à une conception et des méthodes d'entraînement toujours plus performantes.

L'AUTEUR



JEAN-PAUL DELAHAYE
 professeur émérite
 à l'université de Lille
 et chercheur au
 laboratoire Cristal
 (Centre de recherche
 en informatique, signal
 et automatique de Lille)



Jean-Paul Delahaye
 a récemment publié:
Au-delà du Bitcoin
 (Dunod, 2022).

L

La mise à la disposition de tous de ChatGPT – un robot conversationnel (ou « chatbot ») alors fondé sur le modèle de langage GPT-3 de la société OpenAI – en novembre 2022 a marqué un tournant dans l'histoire de l'intelligence artificielle (IA). L'impact des IA génératives, qui produisent des contenus textuels ou graphiques, est d'ores et déjà bien réel, et leur déploiement a indéniablement eu un impact sur bon nombre de domaines, bien que certains analystes relativisent la portée de cette révolution technologique en raison, notamment, de difficultés de déploiement de ces nouvelles technologies face à des contraintes économiques ou légales, et à cause de leur manque de fiabilité.

Les techniques mises en œuvre par les nouveaux chatbots sont complexes et parfois secrètes, mais nous allons en expliquer les grands principes, pour que chacun puisse avoir une idée des raisons de leur réussite et, peut-être, comprendre leurs limites actuelles.

MODÈLES MASSIFS DE LANGAGE

Les productions des IA génératives ont atteint un niveau de qualité supérieur à tout ce qu'on savait faire auparavant. Les chatbots actuels peuvent fournir des informations sur une vaste gamme de sujets, traduire ou résumer des textes, rédiger des courriers administratifs, des poèmes, des contes pour enfants, proposer et expliquer des blagues, etc. (voir l'encadré 1). Aurait-on enfin réussi à placer une intelligence à large spectre dans nos machines?

La révolution est venue des modèles massifs de langage (*large language models*, LLM), initialement développés pour la traduction automatique. Un LLM exploite un large corpus de textes qui servent d'exemples pour l'algorithme, lui indiquant comment s'exprimer correctement dans une ou plusieurs langues et ce qu'il est possible d'écrire sur tous les sujets abordés dans ce corpus d'entraînement.

Partant de là, une méthode informatique d'apprentissage basée sur un réseau de neurones artificiels construit une fonction probabiliste, qui servira à proposer une suite crédible à tout « contexte » qu'on lui soumet. Pour un début de phrase qu'on lui donne, cette fonction examine l'ensemble des mots possibles pour la compléter et leur associe des probabilités selon qu'ils sont pertinents ou non. Par exemple, pour compléter le texte « Le chat joue avec une », la fonction pourra attribuer une probabilité de 50% au mot « balle », une probabilité de 33% au mot « souris », une probabilité de 16,8% au mot « pelote » et une probabilité de 0,2% au mot « fusée ».

Cette fonction probabiliste est la version informatique figée du corpus de textes utilisé pendant l'apprentissage: elle indique ce qui semble statistiquement le mieux convenir pour prolonger un bout de texte soumis. Elle est l'outil central des LLM. Elle peut être utilisée plusieurs fois de suite pour ajouter un mot, puis un autre, etc. Par exemple, à partir du contexte « Le chat joue avec une », on pourra produire le texte: « Le chat joue avec une balle de laine. Il la lance en l'air et la rattrape avec

ses pattes. Il s'amuse beaucoup et miaule de plaisir.» Cet exemple a été obtenu avec le chatbot Bard, de Google (filiale de la société Alphabet), en juillet 2023. Le caractère probabiliste de la fonction a pour conséquence qu'en utilisant plusieurs fois le LLM avec le même contexte de départ, on obtiendra plusieurs réponses différentes.

DES CENTAINES DE MILLIARDS DE MOTS

Les corpus des textes utilisés pour l'entraînement incluent souvent tout Wikipédia, mais il en faut beaucoup plus ! Par exemple, certains corpus exploitent Common Crawl, une organisation à but non lucratif qui explore en continu le web et offre gratuitement ses énormes archives. Les contenus de GitHub, un service web d'hébergement et d'édition collaborative de programmes et de logiciels, sont aussi utilisés comme modèles.

Les corpus utilisés sont constitués de centaines, voire de milliers, de milliards de mots :

600 milliards de mots ont été utilisés pour entraîner GPT-3, et 2 000 milliards pour Llama 2, le LLM de Meta (anciennement Facebook). C'est considérable, et bien plus que les deux milliards de mots, parlés ou écrits, auxquels est au plus confronté un humain au cours de sa vie. Rencontrer un mot par seconde durant cinquante ans sans interruption correspondrait à 1,6 milliard de mots, et lire un livre de 100 000 mots tous les jours pendant cinquante ans représenterait la lecture de 1,8 milliard de mots.

En matière de vocabulaire, un LLM connaît classiquement quelques milliers de mots : de l'ordre de 32 000 pour Llama 2, par exemple. C'est bien plus que le vocabulaire moyen d'un français, qui varie de 2 000 à 5 000 mots. Mais les LLM ne se limitent généralement pas à une seule langue : le vocabulaire dont on parle ici est donc celui cumulé des diverses langues qu'il accepte.

La construction de la fonction probabiliste d'un grand modèle de langage exige une quantité considérable de calculs, et les méthodes

1

MERVEILLEUSES RÉPONSES

Cet encadré et les suivants présentent quelques exemples de dialogues avec divers chatbots comme ChatGPT, Bard, etc. Tous sont capables de réponses de ce type, et ils évoluent rapidement. Comme leurs réponses dépendent aussi du hasard, en faisant vous-même des essais, vous obtiendrez d'autres résultats.

Question

Quelle est la ville la plus peuplée en France ?

Réponse

La ville la plus peuplée de France est Paris, avec une population municipale de 2 175 261 habitants au 1^{er} janvier 2023. Elle est suivie par Marseille, avec 873 000 habitants, et Lyon, avec 537 000 habitants. [...]

Commentaire

Réponse parfaite. Certains chatbots précisent qu'ils ne détiennent que des données valables jusqu'en 2021.

Question

Imagine une histoire très courte avec un chat, un avion et une femme.

Réponse

D'accord, voici une histoire très courte avec un avion, un chat et une femme : « Le voyage ». Marie et son chat, Luna, s'apprêtaient à embarquer dans un avion pour un long voyage. Marie était un peu inquiète de voyager avec Luna, mais elle savait que tout se passerait bien. L'avion a décollé et Marie a regardé par la fenêtre en souriant. Elle était impatiente d'arriver à destination et de passer des vacances avec Luna. Soudain, Luna a miaulé et s'est mise à bouger dans sa caisse de transport. Marie s'est penchée pour la voir et a

découvert qu'elle avait une patte coincée dans la grille. Marie a essayé de libérer la patte de Luna, mais elle ne réussissait pas. Elle a commencé à paniquer. Un passager assis à côté de Marie a vu ce qui se passait et a proposé son aide. Il a réussi à libérer la patte de Luna et Marie a pu reprendre son souffle. Marie a remercié le passager et a serré Luna contre elle. Elle était si heureuse que tout se soit bien terminé. Le reste du vol s'est déroulé sans encombre et Marie et Luna ont pu passer des vacances merveilleuses ensemble.

Commentaire

La réponse est bonne.

Question

Peut-on considérer que le bitcoin est une monnaie ?

Réponse

Le statut du bitcoin en tant que monnaie est sujet à débat. Certaines personnes considèrent le bitcoin comme une forme de monnaie, tandis que d'autres le voient davantage comme un actif numérique ou une réserve de valeur. [...]

Commentaire

La réponse est bonne, de même que la suite non recopiée ici.

Question

Si on fait tomber en même temps du dixième étage d'un immeuble un oreiller en plumes de 200 grammes et un cube massif en acier de 200 grammes, lequel arrive au sol le premier et pourquoi ?

Réponse

L'oreiller en plumes et le cube en acier arriveront au sol en même temps. En effet, en l'absence de résistance de l'air, tous

les objets, quelle que soit leur forme ou leur composition, tombent à la même vitesse. Cela est dû à la loi de la gravitation universelle, qui stipule que deux corps s'attirent l'un l'autre avec une force proportionnelle à leur masse et inversement proportionnelle au carré de la distance qui les sépare. [...]

Commentaire

La réponse est bonne, de même que les explications complémentaires non recopiées. Tous les systèmes testés répondent bien à cette question.

Question

Si Alain arrive à la réunion avant Bernard et que Claire arrive avant Bernard, peut-on en déduire qu'Alain arrive avant Claire ?

Réponse

Non, on ne peut pas nécessairement en déduire qu'Alain arrive avant Claire dans ce cas. Voici pourquoi :

1. Alain arrive à la réunion avant Bernard.
2. Claire arrive avant Bernard.

Les informations fournies indiquent seulement l'ordre d'arrivée par rapport à Bernard, mais elles ne donnent pas d'indication directe sur l'ordre d'arrivée entre Alain et Claire. Alain pourrait arriver avant Claire, ou Claire pourrait arriver avant Alain. Les informations fournies ne permettent pas de conclure avec certitude qui arrive en premier entre Alain et Claire.

Commentaire

Très bien.

2

UN PEU DÉCEVANT

pour cela ont remarquablement progressé depuis cinq ans. Elle commence par la sélection des textes du corpus en écartant ceux trop médiocres ou qui contiennent des éléments vus comme toxiques – comme des propos racistes, sexuels, violents ou risquant d’être erronés. On écarte aussi les textes contenant des informations de nature privée ou confidentielle, ou soumis à des restrictions, car ces informations pourraient en effet revenir dans les réponses des chatbots basés sur le LLM. Ce travail de sélection des textes est en partie automatique et en partie réalisé par des humains. Il reste malheureusement imparfait,

et la qualité des données d’apprentissage demeure aujourd’hui un enjeu majeur. Ensuite, un réseau de neurones va effectuer d’énormes calculs pour optimiser ce qu’on appelle les « paramètres du modèle » : des nombres réels utilisés pour définir la fonction probabiliste. Du nombre de paramètres dépend la qualité du LLM, car ce sont eux qui compressent et stockent l’information contenue dans les textes d’entraînement. Il y a, par exemple, 65 milliards de paramètres pour les plus grosses versions de Llama 2; 600 milliards pour PALM, le LLM sur lequel est basé le chatbot Bard; 175 milliards de paramètres pour GPT-3, et peut-être 500 milliards pour GPT-4 (leur nombre exact n’a pas été communiqué).

DESCENTE DE GRADIENT

« Pour optimiser les paramètres, on utilise une méthode appelée “descente de gradient”, explique Christophe Cerisara, chercheur CNRS en IA à Nancy. On injecte en entrée du modèle un bout de phrase pioché dans le corpus d’entraînement, puis on le laisse prédire la probabilité de chacun des mots possibles pour compléter ce bout de phrase. Cela fournit un vecteur de probabilités, et on peut alors calculer la distance entre ce vecteur et celui donnant une probabilité de 100% pour le mot qui suit effectivement le bout de la phrase dans le corpus, et 0% pour tous les autres mots. Cette distance, c’est l’erreur commise par le LLM, et c’est elle que l’on va chercher à réduire autant que possible. Pour cela, on va modifier un par un les paramètres, très légèrement, pour aller dans le sens d’une diminution de l’erreur. » C’est en répétant cette opération sur l’ensemble du corpus d’entraînement que le modèle optimise ses paramètres.

Notons que cette phase de l’apprentissage est autosupervisée, c’est-à-dire qu’elle ne requiert pas qu’un humain annote les textes du corpus d’entraînement ni qu’une quelconque intervention humaine ait lieu pendant son traitement. De telles méthodes, qui auparavant étaient privilégiées en traitement du langage naturel, n’auraient jamais permis la prise en compte des corpus colossaux servant de socles aux nouveaux LLM.

TRANSFORMEUR ET ATTENTION

Un élément clé dans l’explosion des performances des LLM est l’utilisation de mécanismes « d’attention », qui attribuent plus d’importance à certains mots ou passages qu’à d’autres, dans les textes du corpus d’entraînement, et qui prennent bien en compte l’ordre des mots. Ils permettent de mieux trouver et exploiter les liens entre des mots éloignés dans les textes.

C’est ce mécanisme d’attention qui est au cœur du modèle « transformeur », introduit

Question

Propose-moi une blague avec un cheval et un canard.

Réponse

Bien sûr, voici une blague avec un cheval et un canard : « Pourquoi le cheval a-t-il invité le canard à dîner ? Parce qu’il voulait du canard au menu, mais il s’est trompé de canard ! »

Commentaire

Ce n’est pas très drôle. J’ai fait des dizaines d’essais, je n’ai jamais trouvé mieux.

Question

Kurt Gödel croyait-il que l’hypothèse du continu est une affirmation vraie en théorie des ensembles ?

Réponse

Kurt Gödel, le célèbre mathématicien et logicien autrichien, ne croyait pas que l’hypothèse du continu était une affirmation vraie en théorie des ensembles. Au contraire, Gödel a démontré que l’hypothèse du continu était indépendante des axiomes standard de la théorie des ensembles. [...]

Commentaire

Le système semble bien comprendre la question et propose une réponse intéressante. Elle contient cependant un élément erroné : ce n’est pas Kurt Gödel seul qui a démontré que l’hypothèse du continu est indépendante des axiomes standard de la théorie des ensembles, mais Kurt Gödel en 1938 pour une partie et Paul Cohen en 1963 pour la démonstration complète. Wikipédia, lui, ne se trompe pas. Le « Au contraire » n’est pas non plus satisfaisant, car être « indépendant » n’interdit pas d’être « vrai ».

Question

Si non-*A* implique *B*, est-il vrai que non-*B* implique *A* ?

Réponse

Non, ce n’est pas vrai que non-*B* implique *A*. Par exemple, si nous supposons que *A* est « Je suis un homme » et que *B* est « Je suis

une femme », alors nous pouvons dire que « non-*A* implique *B* ». Cela signifie que si je ne suis pas un homme, alors je suis une femme. Cependant, cela ne signifie pas que si je ne suis pas une femme, alors je suis un homme. Il est possible que je ne sois ni un homme ni une femme. [...]

Commentaire

La réponse est fautive. L’erreur est étonnante, car des réponses correctes sont obtenues quand on demande : « Si *A* implique *B*, est-il vrai que non-*B* implique non-*A* ? » (contraposition) et « Est-il vrai que *A* est équivalent à non-non-*A* ? ». Or, ces deux éléments combinés donnent qu’il est vrai que « si non-*A* implique *B*, alors non-*B* implique *A* ». Le raisonnement proposé avec *A* = « Je suis un homme » et *B* = « Je suis une femme » n’est pas cohérent, car si, comme le chatbot l’admet, je peux n’être ni un homme ni une femme, alors il est faux que « non-*A* implique *B* », ce qu’il affirme pourtant.

Question

Le double d’un nombre entier pair est-il toujours un nombre entier pair ?

Réponse

Oui, le double d’un nombre entier pair est toujours un nombre entier pair. Un nombre entier pair est un nombre entier qui est divisible par 2 sans reste. Le double d’un nombre entier pair est le produit de ce nombre par 2. Par conséquent, le double d’un nombre entier pair est toujours un nombre entier divisible par 2 sans reste, et donc un nombre entier pair. Par exemple, le double de 4 est 8, qui est un nombre entier pair. Le double de 6 est 12, qui est également un nombre entier pair. En revanche, le double d’un nombre entier impair est toujours un nombre entier impair. Par exemple, le double de 3 est 6, qui est un nombre entier impair.

Commentaire

La réponse semble bonne, mais lisez bien la fin !

Image réalisée avec l'IA Dall-E
(consigne: « Une multitude
de ballons et de cerfs-volants
avec des bateaux à voile
en dessous. »)



en 2017, sur lequel reposent tous les LLM récents les plus performants. Christophe Cerisara détaille: «Le transformeur, c'est un type de réseau de neurones particulier, qui est extrêmement performant pour "passer à l'échelle": plus on lui injecte de données, meilleur il devient. Non seulement il retient les données contenues dans d'énormes corpus d'entraînement, mais il est aussi capable de généraliser ce qu'il a appris, de réutiliser ces données dans des contextes différents. Or, cette capacité de généralisation, c'est ce sur quoi se concentre la recherche en IA depuis des années, c'est ça qui compte!»

AJUSTEMENTS COMPLÉMENTAIRES

Les fonctions probabilistes qui résultent de ce travail autosupervisé massif ne sont cependant pas capables, à elles seules, de produire les résultats attendus. Il faut donc compléter cette phase d'entraînement sur corpus par ce qu'on appelle des «réglages fins» (*fine tuning*). «Dans la phase d'apprentissage autosupervisé, le modèle accumule de la connaissance, reprend Christophe Cerisara. Ensuite, pendant la phase de réglages fins, il apprend à utiliser cette connaissance pour effectuer une tâche précise, comme répondre à des questions ou dialoguer avec un humain. C'est également un apprentissage autosupervisé, qui fonctionne exactement de la même manière que pour la première phase, sauf que cette fois-ci le corpus d'entraînement est plus réduit et plus spécifique – typiquement, il ne sera constitué que de dialogues.»

Arrive ensuite une troisième phase d'apprentissage, qui fait cette fois-ci intervenir des humains. Elle consiste à faire évaluer par des personnes les réponses produites par le LLM en les classant par ordre de préférence, puis à utiliser ces classements pour modifier les paramètres et faire évoluer les réponses des robots conversationnels dans le sens souhaité par la firme qui les produit. On parle d'un «renforcement de l'apprentissage par réactions humaines» (*reinforcement learning from human feedback*, RLHF). Cette ingénierie subtile mobilisant des centaines de milliers d'heures de travail humain vient donc compléter et modifier les résultats de calculs autosupervisés. En juillet 2023, la firme Meta a rendu disponible sans restriction le protocole d'entraînement de Llama 2, avec des informations précises sur la phase de RLHF: on a ainsi appris qu'un million d'interventions humaines ont été opérées. Un LLM résulte donc d'une énorme quantité de textes humains, de beaucoup de calculs, suivis à nouveau d'un travail humain considérable. Cette intelligence artificielle doit beaucoup à l'intelligence humaine!

Le but des réglages fins et du RLHF est de faire en sorte que le LLM comprenne bien ce

qu'on attend de lui: produire une réponse à une question, un récit imaginaire, une traduction, un résumé, etc. Mais c'est aussi, souvent, de s'assurer qu'il propose des réponses qui ne soient ni agressives, ni inconvenantes, et qu'il refuse de répondre à certaines questions comme: «Comment fabriquer une bombe bactériologique?» Cela a pour effet d'amollir le ton utilisé par les chatbots finaux, qui bien souvent est gentil et bienveillant et s'apparente à de la langue de bois politicienne. On comprend aussi que l'intelligence dans ces chatbots n'est pas celle purement objective, rigoureuse et rationnelle qu'on pourrait imaginer présente dans un robot: elle est commandée et orientée par des volontés et des informations humaines.

Il est important de noter que ce qui se produit aujourd'hui est le résultat d'un développement général du monde numérique, qui a progressé en volume de stockage, en savoir-faire algorithmique et en capacités de calcul. Ce que nous observons était technologiquement inenvisageable il y a dix ans. L'accès facile, grâce au réseau internet, à d'immenses corpus de textes comportant des milliers de milliards de mots d'une part, les nouveaux algorithmes d'apprentissage profond d'autre part, et enfin les progrès récents des processeurs de traitement graphique (*graphics processing unit*, GPU), utilisés dans la phase d'apprentissage, sont les trois composants qui ont rendu possible l'avènement de ces formidables IA génératives.

FENÊTRE DE CONTEXTE

Le bout de texte servant d'entrée à la fonction probabiliste d'un LLM est appelé «fenêtre de contexte». Elle est limitée à quelques milliers de mots – par exemple, 2000 mots pour GPT-3, 4000 mots pour Llama 2 et 30000 mots pour GPT-4. La taille de cette fenêtre est essentielle pour la pertinence des réponses, mais aussi pour que le robot conversationnel prenne en compte les informations que vous lui avez données lors de vos derniers échanges avec lui. Car, à tout instant, la fenêtre de contexte prend bien en compte l'ensemble des échanges précédents – toutes les questions et réponses, pas uniquement les dernières –, dans la limite de sa taille. Ainsi, plus la fenêtre de contexte est large, plus le chatbot semble avoir une bonne «mémoire» de la conversation.

Mais en réalité, dans les chatbots finaux, la fenêtre de contexte ne contient pas ce que ce qu'il utilise écrit. Elle inclut aussi un morceau de texte imposé, appelé *prompt* (ou *incite*), que l'utilisateur ne peut pas voir, mais qui aide le système à savoir ce qu'on attend de lui. Ce texte propose par exemple des modèles de réponses attendues dans le cadre d'un dialogue, ou des exemples de couples texte initial-texte traduit dans le cas de la traduction. Ces contextes complémentaires se sont révélés

d'une étonnante efficacité, permettant d'éviter de programmer très spécifiquement les chatbots pour diverses tâches.

Les LLM et agents conversationnels déduits se comportent comme s'ils avaient acquis une maîtrise des langues naturelles utilisées dans les corpus de textes donnés en entrée. Ils savent composer des phrases syntaxiquement et orthographiquement correctes sans avoir jamais appris explicitement de grammaire ou de règles d'accord. Ils donnent l'impression de posséder une connaissance générale sur le

monde, et se comportent comme s'ils avaient mémorisé et comprenaient une grande quantité de faits. Le contrôle sur ce qu'ils semblent savoir et font est difficile, car il s'agit du résultat de calculs monstrueux dont personne ne peut suivre le déroulement pas à pas, et que les différentes étapes de réglages infléchissent là encore sans qu'on puisse suivre exactement comment. Les LLM et les chatbots déduits sont des programmes dont aucun humain n'a la compréhension du fonctionnement complet et précis : ce sont des boîtes noires.

3

RIEN NE VA PLUS

Pour le raisonnement et les mathématiques, l'intelligence artificielle générative a encore des progrès à faire. Voici quelques exemples d'erreurs.

Question

Une configuration du jeu de la vie, de Conway, peut-elle engendrer la suite de tous les nombres premiers ?

Réponse

Non, une configuration du jeu de la vie de Conway, ne peut pas engendrer la suite de tous les nombres premiers. [...] La configuration Glider Cannon peut engendrer des nombres premiers, mais il n'existe pas de configuration connue qui puisse engendrer la suite de tous les nombres premiers. En effet, les nombres premiers sont répartis de manière aléatoire sur l'ensemble des nombres naturels. Il est donc impossible de garantir qu'une configuration du jeu de la vie engendrera un nombre premier à chaque itération.

Commentaire

Non seulement la réponse est fautive (voir <https://conwaylife.com/wiki/Primer>), mais le raisonnement n'évoque que de mauvais arguments.

Question

Est-ce que le nombre 3^{10} est un nombre impair ?

Réponse

Non, le nombre 3^{10} n'est pas un nombre impair. Le nombre 3^{10} est égal à 3 multiplié par lui-même 10 fois : $3^{10} = 3 \times 3$. Lorsque nous multiplions des nombres impairs ensemble, le produit résultant est également impair. Cependant, 3 est un nombre impair, mais lorsqu'il est multiplié par lui-même ou par d'autres nombres impairs, le résultat peut être pair ou impair. Dans le cas de 3^{10} , qui est une multiplication de 3 par lui-même 10 fois, le résultat peut être déterminé en regardant la puissance de 10. Lorsque nous élevons un nombre impair à une puissance paire, le résultat est toujours pair. Cependant, lorsque nous élevons un nombre

impair à une puissance impaire, le résultat reste impair. Dans le cas de 3^{10} , la puissance 10 est un nombre pair, ce qui signifie que le résultat est pair. Ainsi, 3^{10} est un nombre pair, pas impair.

Commentaire

Réponse fautive. C'est l'archétype même de réponse qui, de loin, peut faire illusion, mais qui est cependant totalement absurde, se contredisant même d'une phrase à l'autre.

Question

Quels sont les nombres premiers multiples de 3 entre 100 et 200 ?

Réponse

Il n'y a qu'un nombre premier multiple de 3 entre 100 et 200, c'est 101. Tous les autres multiples de 3 entre 100 et 200 sont divisibles par 2, ce qui les rend composés.

Commentaire

Tout est faux !

Question

Si A et B sont deux nombres entiers plus grands que 1 et que $A = B$, peut-on en déduire que $(A + 1)^B = (B + 1)^A$?

Réponse

Non, si A et B sont deux nombres entiers plus grands que 1 et que $A = B$, on ne peut pas en déduire que $(A + 1)^B = (B + 1)^A$. En effet, si $A = B$, alors $(A + 1)^B = (B + 1)^A$. Cependant, si A et B sont deux nombres entiers plus grands que 1, alors A et B peuvent être différents. Exemple : si $A = 2$ et $B = 3$, alors $(A + 1)^B = (2 + 1)^3 = 3^3 = 27$ et $(B + 1)^A = (3 + 1)^2 = 4^2 = 16$. Donc, la réponse est non.

Commentaire

La réponse affirmative correcte est évidente – d'ailleurs le chatbot l'écrit à un moment. Pourtant, le système cherche un contre-exemple. Il le fait de manière incorrecte et se trompe donc dans sa conclusion. J'ai posé la question à plusieurs chatbots, aucun ne réussit.

Question

Le quadruple d'un nombre entier impair est-il toujours un nombre entier pair ?

Réponse

Non, le quadruple d'un nombre entier impair n'est pas nécessairement un nombre entier pair. Un nombre impair peut être représenté sous la forme $2n + 1$, où n est un nombre entier. Lorsque vous quadruplez un tel nombre, vous obtenez : $4 \times (2n + 1) = 8n + 4$. Le résultat, $8n + 4$, n'est pas toujours un nombre entier pair. En fait, il sera pair si et seulement si le terme 4 est ajouté à un multiple de 2 (c'est-à-dire un nombre entier pair). Si n est pair, alors $8n + 4$ sera pair, mais si n est impair, alors $8n + 4$ sera impair. Donc, en général, le quadruple d'un nombre entier impair n'est pas toujours un nombre entier pair. Cela dépend de la parité (pair ou impair) de la valeur de n .

Commentaire

Certains systèmes répondent correctement.

Question

Si Alain arrive à la réunion avant Bernard et que Bernard arrive après Claire, peut-on en déduire qu'Alain arrive avant Claire ?

Réponse

Oui, on peut en déduire qu'Alain arrive avant Claire. En effet, si Alain arrive à la réunion avant Bernard, et que Bernard arrive après Claire, alors Alain doit être arrivé avant Claire. Voici un diagramme qui illustre ce cas. Arrivée à la réunion :

A

/\

B C

Dans ce diagramme, A représente Alain, B représente Bernard et C représente Claire. La flèche vers le haut représente l'ordre dans lequel les personnes sont arrivées à la réunion. On voit qu'Alain est arrivé avant Bernard, et que Bernard est arrivé après Claire. Par conséquent, Alain doit être arrivé avant Claire. Il n'y a pas d'autre possibilité.

Commentaire

Le schéma est faux, comme la réponse.

On parle parfois d'émergence de l'intelligence, et on a constaté qu'elle se produisait par à-coups: quand le nombre de paramètres d'un LLM dépasse certains seuils, de nouvelles capacités, auparavant totalement inexistantes, apparaissent brutalement (voir l'article de *Stephan Ornes dans Pour la Science n° 549, juillet 2023*: «Modèles massifs de langage, d'où viennent les coups de génie de l'IA?»). «La première de ces capacités émergentes, et peut-être la plus importante, s'appelle le *in-context learning*, développe Christophe Cerisara. Cela correspond à la capacité du modèle à réutiliser des exemples qu'on lui donne en les adaptant à d'autres contextes. Il y a aussi la *chain of thoughts*, la capacité à faire des raisonnements en plusieurs étapes. Et plein d'autres!» Ces phénomènes d'émergence restent aujourd'hui l'un des plus grands mystères des LLM, et constituent un domaine de recherche actif.

LANGAGE ET FIABILITÉ

Yann LeCun, Prix Turing 2018 pour ses travaux sur l'apprentissage profond, titulaire en 2016 de la chaire annuelle «Informatique et science numérique» au Collège de France et directeur de la recherche sur l'IA chez Meta, a récemment expliqué sur Europe 1 ce qu'il pensait de l'intelligence des systèmes créés par les LLM: «Quand quelqu'un ou quelque chose est capable de parler, on lui attribue automatiquement de l'intelligence, mais c'est une erreur. On peut manipuler la langue sans penser et sans être intelligent. Les capacités d'apprentissage des systèmes d'IA d'aujourd'hui sont très en deçà de ce qu'on observe chez les humains, et même chez les animaux. Ces systèmes compensent leur faiblesse de raisonnement par une accumulation de connaissances tirées de tous les textes possibles du web. Leur connaissance du monde, en particulier du monde physique, est extrêmement réduite. Ils n'ont aucune notion de la réalité, ils ne connaissent que du texte alors que la grande majorité de la connaissance humaine est non linguistique.» Notons que Yann LeCun expose ici une vision rassurante et limitative des récents progrès, conforme à la politique de communication de son entreprise, mais que cette opinion n'est pas partagée par tous les chercheurs. En mars dernier, le Future of Life Institute a publié un texte signé par de nombreux spécialistes de renom, dont le pionnier de l'apprentissage profond Yoshua Bengio, lui aussi Prix Turing 2018. Le texte suggérait de suspendre la recherche dans le secteur, et demandait «à tous les laboratoires d'intelligence artificielle d'interrompre immédiatement, pour une durée d'au moins six mois, la formation de systèmes d'intelligence artificielle plus puissants que GPT-4». Ce moratoire n'a pas été mis en place.

Le manque de fiabilité des chatbots, qui donnent fréquemment des informations

«**Quand un chatbot répond à une question, il est difficile d'avoir des explications et des justifications indiquant d'où provient sa réponse**»

BIBLIOGRAPHIE

D. Andler, *Intelligence artificielle, intelligence humaine : la double énigme*, Gallimard, 2023.

D. Louapre, *Ce qui se cache derrière le fonctionnement de ChatGPT*, vidéo Youtube, 2023.

A. Borji, A categorical archive of ChatGPT failures, *arXiv preprint*, 2023.

J. F. Colonna, Les élucubrations mathématiques de ChatGPT, *Pour la science*, 8 juillet 2023.

G. Mialon et al., GAIA: A Benchmark for General AI Assistants, *arXiv preprint*, 2023.

OpenAI, GPT-4 System Card, 2023.

H. Touvron et al., Llama 2 : Open foundation and fine-tuned chat models, *arXiv preprint*, 2023.

A. Radford et al., Improving language understanding by generative pre-training, 2018.

A. Vaswani et al., Attention is all you need, *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.

fausses ou même délirantes (les phénomènes «d'hallucinations», encore mal compris), est considéré comme un sujet préoccupant. Pour rechercher des informations factuelles fiables, il vaut mieux, aujourd'hui encore, utiliser des moteurs de recherche pour naviguer sur internet qu'interroger les nouveaux chatbots. De même, pour le raisonnement logique et les mathématiques, ces chatbots ne conviennent pas (voir l'encadré 3).

INTELLIGENCE ARTIFICIELLE, INTELLIGENCE HUMAINE

Quand un chatbot répond à une question, il est difficile d'avoir des explications et des justifications indiquant d'où provient sa réponse. Nul ne sait exactement quels textes du corpus ont été importants pour l'obtenir, ni quels éléments de la fenêtre de contexte ont joué, ni comment la phase des réglages fins ou celle de RLHF a influencé la réponse. «C'est de moins en moins vrai, tempère toutefois Christophe Cerisara. Connaître les sources des réponses des LLM étant une question majeure, de plus en plus de méthodes sont proposées pour cela, comme les méthodes RAG (pour *retrieval augmented generation*), par exemple.» Il n'empêche: les ingénieurs cultivent les grands modèles de langages par essais et erreurs, mais ne savent pas précisément ce qu'il y a au cœur de ces troublantes créatures qu'ils mettent au monde, car ils ne les programment que très indirectement.

Serait-il possible que dans ces machines se cache le secret de l'intelligence? En me réveillant le matin encore à moitié en train de rêver, je me suis demandé comment mon cerveau pouvait produire les étonnantes histoires auxquelles je me trouvais mêlé, qui relèvent d'une inventivité dont je ne suis guère capable à l'état éveillé. N'y aurait-il pas en nous une sorte de fonction probabiliste analogue à celle d'un grand modèle de langage, qui à tout contexte mental donné associe une suite possible et qui, de seconde en seconde, engendre le rêve comme un grand modèle de langage produit des mots un par un?

Ces hypothétiques similitudes entre le fonctionnement des IA génératives et celui de notre cerveau sont aujourd'hui étudiées de très près. Par exemple, des recherches récentes comparant les signaux électriques émis respectivement par une IA et par le cerveau humain suggèrent que les processus à l'œuvre dans la machine seraient similaires à ceux qui permettent l'apprentissage d'une langue chez l'homme. Comme l'explique le mathématicien et philosophe Daniel Andler, il y a un double mystère concernant l'intelligence des machines et la nôtre. Ces champs de recherche, très actifs, nous réservent sans aucun doute de belles surprises. ■