www.quantamagazine.org /how-quickly-do-large-language-models-learn-unexpected-skills-20240213/

À quelle vitesse les grands modèles de langage apprennent-ils des compétences inattendues
Quanta Magazine

Une nouvelle étude suggère que les capacités dites émergentes se développent progressivement et de manière prévisible, selon la façon dont vous les mesurez.

By Stephen Ornes :

Il y a deux ans, dans le cadre d'un projet appelé Beyond the Imitation Game, ou BIG-bench, 450 chercheurs ont compilé une liste de 204 tâches conçues pour tester les capacités des grands modèles de langage, qui alimentent les chatbots comme ChatGPT. Sur la plupart des tâches, les performances se sont améliorées de manière prévisible et fluide au fur et à mesure que les modèles évoluaient : plus le modèle était grand, plus il s'améliorait.
Mais avec d'autres tâches, le saut de capacité n'a pas été facile.
La performance est restée proche de zéro pendant un certain temps, puis les performances ont bondi.
D'autres études ont trouvé des sauts similaires dans les capacités.

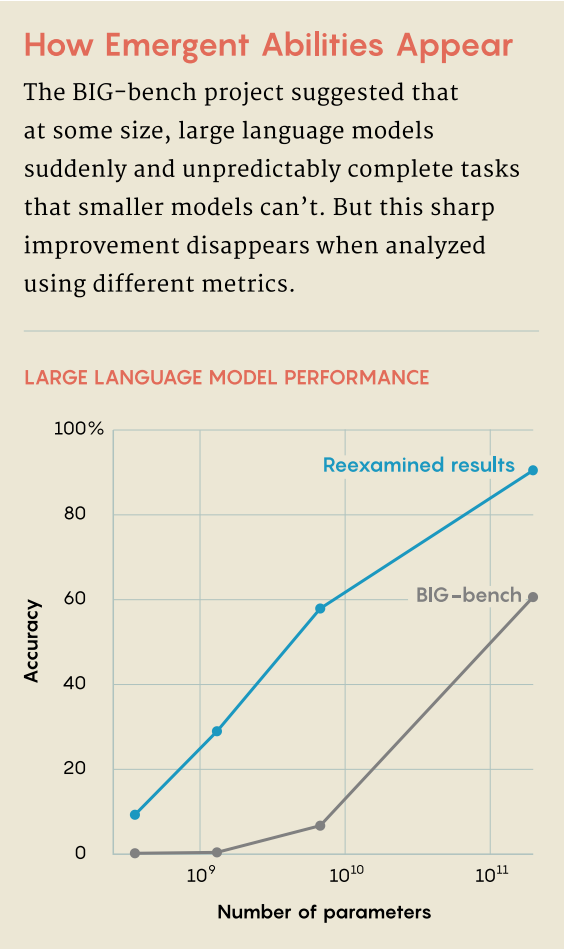
Les auteurs ont décrit cela comme un comportement « révolutionnaire »; d'autres chercheurs l'ont comparé à une transition de phase en physique, comme lorsque l'eau liquide gèle en glace.
Dans un article publié en août 2022, les chercheurs ont noté que ces comportements sont non seulement surprenants, mais aussi imprévisibles, et qu'ils devraient éclairer l'évolution des conversations sur la sécurité,

le potentiel et les risques de l'IA.
Ils ont appelé les capacités « émergentes », un mot qui décrit les comportements collectifs qui n'apparaissent qu'une fois qu'un système atteint un niveau élevé de complexité.

Mais les choses ne sont peut-être pas si simples.
Un nouvel article d'un trio de chercheurs de l'Université de Stanford postule que l'apparition soudaine de ces capacités n'est qu'une conséquence de la façon dont les chercheurs mesurent les performances du LLM.
Les capacités, disent-ils, ne sont ni imprévisibles ni soudaines.
« La transition est beaucoup plus prévisible que ce que les gens lui attribuent », a déclaré Sanmi Koyejo, informaticien à Stanford et auteur principal de l'article.
« Les fortes affirmations d'émergence ont autant à voir avec la façon dont nous choisissons de mesurer qu'avec ce que font les modèles. »

Ce n'est que maintenant que nous voyons et étudions ce comportement en raison de la taille de ces modèles.
Les grands modèles linguistiques s'entraînent en analysant d'énormes ensembles de données de texte – des mots provenant de sources en ligne, notamment des livres, des recherches sur le Web et Wikipédia – et en trouvant des liens entre des mots qui apparaissent souvent ensemble.
La taille est mesurée en termes de paramètres, à peu près analogues à toutes les façons dont les mots peuvent être connectés.
Plus il y a de paramètres, plus un LLM peut trouver de connexions. GPT-2 avait 1,5 milliard de paramètres, tandis que GPT-3.5, le LLM qui alimente ChatGPT, en utilise 350 milliards. GPT-4, qui a fait ses débuts en mars 2023 et sous-tend désormais Microsoft Copilot, utiliserait 1,75 billion de dollars.

Cette croissance rapide a entraîné une augmentation étonnante des performances et de l'efficacité, et personne ne conteste que des LLM suffisamment grands peuvent accomplir des tâches que les modèles plus petits ne peuvent pas, y compris celles pour lesquelles ils n'ont pas été formés.
Le trio de Stanford qui considère l'émergence comme un « mirage » reconnaît que les LLM deviennent plus efficaces à mesure qu'ils se développent; en fait, la complexité accrue des modèles plus grands devrait permettre de s'améliorer sur des problèmes plus difficiles et plus diversifiés.
Mais ils soutiennent que le fait que cette amélioration semble lisse et prévisible ou irrégulière et nette résulte du choix de la métrique – ou même d'une rareté d'exemples de test – plutôt que du fonctionnement interne du modèle.



Merrill Sherman/*Quanta Magazine*

L'addition à trois chiffres en est un exemple.
Dans l'étude BIG-bench de 2022, les chercheurs ont rapporté qu'avec moins de paramètres, GPT-3 et un autre LLM nommé LAMDA n'ont pas réussi à résoudre avec précision les problèmes d'addition.
Cependant, lorsque GPT-3 s'est entraîné à l'aide de 13 milliards de paramètres, sa capacité a changé comme si elle appuyait sur un interrupteur.
Soudainement, il pourrait s'ajouter – et LAMDA pourrait aussi, à 68 milliards de paramètres.
Cela suggère que la capacité d'ajouter émerge à partir d'un certain seuil.

Mais les chercheurs de Stanford soulignent que les LLM n'ont été jugés que sur l'exactitude : soit ils pouvaient le faire parfaitement, soit ils ne le pouvaient pas.
Ainsi, même si un LLM a prédit correctement la plupart des chiffres, il a échoué.
Cela ne semblait pas correct.
Si vous calculez 100 plus 278, alors 376 semble être une réponse beaucoup plus précise que, disons, −9,34.

Au lieu de cela, Koyejo et ses collaborateurs ont testé la même tâche à l'aide d'une métrique qui attribue un crédit partiel.
« Nous pouvons nous demander : dans quelle mesure prédit-il le premier chiffre ?
Puis le second ?
Puis le troisième ? dit-il.

Koyejo attribue l'idée de ce nouveau travail à son étudiant diplômé Rylan Schaeffer, qui, selon lui, a remarqué que la performance d'un LLM semble changer avec la façon dont ses capacités sont mesurées.
Avec Brando Miranda, un autre étudiant diplômé de Stanford, ils ont choisi de nouvelles mesures montrant qu'à mesure que les paramètres augmentaient, les LLM prédisaient une séquence de chiffres de plus en plus correcte en plus des problèmes.
Cela suggère que la capacité d'ajouter n'est pas émergente – ce qui signifie qu'elle subit un saut soudain et imprévisible – mais progressive et prévisible. Ils constatent qu'avec un autre bâton de mesure, l'émergence disparaît.

Mais d'autres scientifiques soulignent que le travail ne dissipe pas complètement la notion d'émergence.
Par exemple, l'article du trio n'explique pas comment prédire quand les mesures, ou lesquelles, montreront une amélioration abrupte dans un LLM, a déclaré Tianshi Li, informaticien à la Northeastern University.
« En ce sens, ces capacités sont toujours imprévisibles », a-t-elle déclaré.
D'autres, comme Jason Wei, un informaticien qui travaille maintenant chez OpenAI et qui a compilé une liste de capacités émergentes et était l'un des auteurs de l'article de BIG-bench, ont fait valoir que les rapports antérieurs sur l'émergence étaient solides parce que pour des capacités comme l'arithmétique, la bonne réponse est vraiment tout ce qui compte.

« Il y a certainement une conversation intéressante à avoir ici », a déclaré Alex Tamkin, chercheur scientifique à la start-up d'IA Anthropic.
Le nouveau document décompose habilement les tâches en plusieurs étapes pour reconnaître les contributions des composants individuels, a-t-il déclaré.
« Mais ce n'est pas tout.
On ne peut pas dire que tous ces sauts sont un mirage.
Je pense toujours que la littérature montre que même lorsque vous avez des prédictions en une seule étape ou que vous utilisez des mesures continues, vous avez toujours des discontinuités, et à mesure que vous augmentez la taille de votre modèle, vous pouvez toujours le voir s'améliorer de manière à sauter.

Et même si l'émergence dans les LLM d'aujourd'hui peut être expliquée par différents outils de mesure, il est probable que ce ne sera pas le cas pour les LLM de demain, plus grands et plus compliqués.
« Lorsque nous ferons passer les LLM au niveau supérieur, ils emprunteront inévitablement des connaissances à d'autres tâches et à d'autres modèles », a déclaré Xia « Ben » Hu, informaticien à l'Université Rice.

Cette réflexion évolutive sur l'émergence n'est pas seulement une question abstraite que les chercheurs doivent se poser.
Pour Tamkin, cela témoigne directement des efforts en cours pour prédire comment les LLM se comporteront.
« Ces technologies sont si vastes et si applicables », a-t-il déclaré.
« J'espère que la communauté utilisera cela comme point de départ pour continuer à souligner l'importance de construire une science de la prédiction pour ces choses. Comment ne pas se laisser surprendre par la prochaine génération de modèles ?

APPARENTÉ:

1. Une nouvelle théorie suggère que les chatbots peuvent comprendre le texte
2. Les capacités imprévisibles qui émergent des grands modèles d'IA
3. Pour enseigner les mathématiques à l'informatique, des chercheurs fusionnent des approches d'IA
4. Les transformateurs vont-ils prendre le contrôle de l'intelligence artificielle ?

Recherche et mise en page par:
Michel Cloutier
CIVBDL
20240216
"C'est ensemble qu'on avance"