



KI wird bald Projekte übernehmen, die Menschen Wochen kosten

Künstliche Intelligenz verbessert sich rasant und könnte bald Projekte übernehmen, für die Menschen Wochen benötigen. Expertenanalysen zeigen, dass führende KI-Modelle im Fortschritt sind und bis 2029 Aufgaben mit menschlicher Expertise in kürzerer Zeit bewältigen könnten.



Die heutigen Systeme der künstlichen Intelligenz (KI) können Menschen bei langen Aufgaben nicht übertreffen, entwickeln sich jedoch **rasant** weiter und könnten die Kluft schneller schließen, als viele erwartet haben, so eine Analyse führender Modelle¹.

Die gemeinnützige Organisation METR aus Berkeley, Kalifornien, entwickelte nahezu 170 echte Aufgaben in den Bereichen Programmierung, Cybersicherheit, allgemeines Denken und maschinelles Lernen und stellte dann eine „menschliche Basislinie“ fest, indem sie die Zeit maß, die Experten benötigten, um diese Aufgaben abzuschließen.

Das Team entwickelte anschließend eine Kennzahl zur Bewertung des Fortschritts von **KI-Modellen**, die "Aufgabenabschluss-Zeithorizont" genannt wird. Dies ist die Zeit, die Programmierer typischerweise benötigen, um die Aufgaben abzuschließen, die KI-Modelle mit einer bestimmten Erfolgsquote erledigen können.

In einem diese Woche auf arXiv veröffentlichten Preprint berichtet METR, dass GPT-2, ein frühes großes Sprachmodell (LLM), das 2019 von OpenAI veröffentlicht wurde, bei allen Aufgaben versagte, die menschliche Experten mehr als eine Minute kosteten. Claude 3.7 Sonnet, das im Februar von dem US-Start-up Anthropic herausgebracht wurde, schloss 50 % der Aufgaben ab, die Menschen 59 Minuten in Anspruch nehmen würden.

Insgesamt hat sich der Zeithorizont der 13 führenden KI-Modelle seit 2019 etwa alle sieben Monate verdoppelt, so die Studie. Das exponentielle Wachstum der KI-Zeithorizonte beschleunigte sich im Jahr 2024, wobei die neuesten Modelle ihren Horizont ungefähr alle drei Monate verdoppeln. Die Arbeit wurde noch nicht formal begutachtet.

Bei dem Fortschritt von 2019 bis 2024 schlägt METR vor, dass KI-Modelle bis 2029 Aufgaben bewältigen können, die Menschen etwa einen Monat benötigen, mit einer Zuverlässigkeit von 50 %, möglicherweise sogar früher.

Ein Monat engagierter menschlicher Expertise, so das Papier, kann ausreichen, um ein neues Unternehmen zu gründen oder wissenschaftliche Entdeckungen zu machen.

Joshua Gans, Professor für Management an der Universität Toronto in Kanada, der über die Wirtschaftlichkeit von KI geschrieben hat, erklärt jedoch, dass solche Vorhersagen nicht besonders nützlich sind. „Extrapolationen sind verlockend, aber es gibt noch so viel, was wir nicht wissen, wie KI tatsächlich eingesetzt wird, damit diese Vorhersagen sinnvoll sind“, sagt er.

Beurteilung von Mensch versus KI

Das Team wählte die Erfolgsquote von 50 %, da sie am robustesten gegenüber kleinen Änderungen in der Datenverteilung war. „Wenn man sehr niedrige oder sehr hohe Schwellenwerte wählt, ändert das Hinzufügen oder Entfernen einer einzigen erfolgreichen oder gescheiterten Aufgabe entsprechend die Schätzung stark“, erklärt Mitautor Lawrence Chan.

Eine Erhöhung der Zuverlässigkeit von 50 % auf 80 % reduzierte den durchschnittlichen Zeithorizont um den Faktor fünf – obwohl die allgemeine Verdopplungszeit und die Trendlinie ähnlich waren.

In den letzten fünf Jahren wurden die Verbesserungen der **allgemeinen Fähigkeiten von LLMs** hauptsächlich durch Skalenerhöhungen – die Menge an Trainingsdaten, Trainingszeit und Anzahl der Modellparameter – vorangetrieben. Das Papier führt den Fortschritt bei der Kennzahl des Zeithorizonts hauptsächlich auf Verbesserungen in der logischen Argumentation, der Nutzung von Werkzeugen, der Fehlerkorrektur und dem Selbstbewusstsein bei der Aufgabenerfüllung zurück.

Der Ansatz von METR, den Zeithorizont zu bewerten, adressiert einige der Einschränkungen bestehender KI-Benchmarks, die nur lose mit der realen Arbeit übereinstimmen und schnell „gesättigt“ werden, wenn sich die Modelle verbessern. Es bietet eine kontinuierliche, intuitive Maßnahme, die bedeutende Fortschritte über einen längeren Zeitraum besser erfasst, so Mitautor Ben West.

Führende KI-Modelle erreichen übermenschliche Leistungen bei vielen **Benchmark-Tests**, hatten jedoch bisher relativ geringe wirtschaftliche Auswirkungen, erklärt West. Die neuesten Forschungen von METR bieten eine teilweise Antwort auf dieses Rätsel: Die besten Modelle zeigen einen Zeitrahmen von etwa

40 Minuten, und es gibt nicht viel wirtschaftlich wertvolle Arbeit, die eine Person in dieser Zeit erledigen kann, so West.

Anton Troynikov, ein KI-Forscher und Unternehmer aus San Francisco, Kalifornien, erklärt jedoch, dass KI einen größeren wirtschaftlichen Einfluss hätte, wenn Organisationen besser bereit wären, zu experimentieren und in eine effektive Nutzung der Modelle zu investieren.

1. Kwa, T. et al. Preprint bei arXiv

<https://doi.org/10.48550/arXiv.2503.14499> (2025).

Referenzen herunterladen

Details

Besuchen Sie uns auf: natur.wiki