

Künstliche Intelligenz:

Mathematiker wollten eine KI überlisten – und scheiterten

Bei einem geheimen Treffen stellten die weltbesten Fachleute schwierigste Mathematikprobleme zusammen. Zu ihrer Überraschung konnten die neuen Sprachmodelle sie lösen. Ist das der Beginn einer neuen Ära?

von [Lyndie Chiou](#)



Eigentlich sollte die KI an den komplexen Fragen scheitern – doch das Gegenteil war der Fall.

Exklusive Übersetzung aus **Scientific American**

Mitte Mai 2025 kamen 30 der weltweit renommiertesten Mathematiker zu einem geheimen Treffen zusammen. Im kalifornischen Berkeley lieferten sie sich ein Kräftemessen mit einem Chatbot. Das Programm sollte Aufgaben lösen, die sich die Experten ausgedacht hatten, um seine mathematischen Fähigkeiten zu testen. Zwei Tage lang konfrontierten die Forschenden die KI mit komplexen Fragen auf Professorenniveau – und wie sie erstaunt feststellten, konnte der Chatbot einige der schwierigsten (lösbaren) Probleme der Welt beantworten. »Meine Kollegen meinten, diese Modelle kämen einem mathematischen Genie nahe«, sagt der Mathematiker Ken Ono von der University of Virginia, der an der Tagung teilnahm.

Der Chatbot, um den es hier geht, wird vom Sprachmodell [o4-mini der US-Firma OpenAI](#) angetrieben. Dieses wurde darauf trainiert, hochkomplexe Schlussfolgerungen zu ziehen. Das Pendant von Google, [Gemini 2.5 Flash](#), verfügt über ähnliche Fähigkeiten. Wie die Sprachmodelle, die hinter früheren Versionen von ChatGPT stecken, lernt auch o4-mini, das nächste Wort in einer Folge von Wörtern vorherzusagen. Im Vergleich zu vorherigen Modellen sind die neuen Versionen wie o4-mini flexibler und schneller – und wurden vermehrt auf spezialisierten Datensätzen mit menschlichem Feedback trainiert. Die so entwickelten Chatbots können somit viel tiefer in komplexe mathematische Probleme eintauchen als bisherige KI-Modelle.

Um die Fähigkeiten von o4-mini zu testen, hat OpenAI [im Jahr 2024 die gemeinnützige Organisation Epoch AI damit beauftragt](#), 300 mathematische Aufgaben auszuarbeiten, deren Lösungen noch nicht veröffentlicht wurden. Zwar können auch die älteren Sprachmodell-Versionen

gewisse Mathematikprobleme beantworten, doch als Epoch AI sie mit den ausgewählten Fragen konfrontierte, lösten selbst die erfolgreichsten unter ihnen weniger als zwei Prozent der Aufgaben. Bei o4-mini sieht das hingegen völlig anders aus.

o4-mini rockt die Matheaufgaben

Epoch AI stellte im September 2024 den frisch promovierten Mathematiker Elliot Glazer ein, um an dem Projekt namens [FrontierMath](#) teilzunehmen. In diesem wurden neue mathematische Fragen mit unterschiedlichen Schwierigkeitsgraden gesammelt: Die ersten drei Stufen deckten jeweils Herausforderungen für Studierende, Hochschulabsolventen und Forschende ab. Im Februar 2025 stellte Glazer fest, dass o4-mini etwa 20 Prozent der Fragen lösen konnte. Deshalb ging er zu einer vierten Stufe über: Er brauchte 100 Fragen, die selbst für Mathematikprofessoren herausfordernd sind. Doch dafür war er auf Hilfe angewiesen, schließlich ist nur eine kleine Gruppe von Menschen auf der Welt in der Lage, solche Fragen auszuarbeiten – geschweige denn sie zu beantworten. Die teilnehmenden Fachleute mussten eine Geheimhaltungsvereinbarung unterzeichnen und durften ausschließlich über die verschlüsselte Messaging-App Signal kommunizieren. Andere Kontaktformen, wie herkömmliche E-Mails, könnten von einem Sprachmodell gescannt und versehentlich für das Training abgegriffen werden, so die Befürchtung.

Die Gruppe machte langsame, aber stetige Fortschritte bei der Suche nach Fragen. Um das Ganze zu beschleunigen, veranstaltete Glazer mit Epoch AI am Wochenende des 17. und 18. Mai 2025 ein geheimes Treffen. Dort sollten die Teilnehmer die letzten zehn Fragen für die ultimative KI-Herausforderung finden. Das Treffen leitete Ono, der die 30 Personen in Sechsergruppen aufteilte. Zwei Tage lang suchten die Akademiker nach Aufgaben, die sie gerade so lösen konnten, die eine KI allerdings vor Probleme stellen würden. Und es gab einen Ansporn: Für jede Aufgabe, die o4-mini nicht lösen konnte, erhielt der Mathematiker, der sie sich ausgedacht hatte, eine Belohnung in Höhe von 7500 US-Dollar.

»Keine Quelle nötig, denn die mysteriöse Zahl wurde von mir
berechnet!«KI-Modell

Am Ende des ersten Tages zeigte sich Ono frustriert über die mangelnden Fortschritte des Teams. »Deshalb habe ich mir ein Problem ausgedacht, von dem jeder auf meinem Gebiet weiß, dass es eine offene Frage in der Zahlentheorie ist – ein gutes Problem auf Doktoranden-Niveau«, sagt er. Er bat o4-mini, die Frage zu beantworten. In den nächsten zehn Minuten konnte Ono fassungslos beobachten, wie der Bot eine Lösung fand und dabei schrittweise seine Schlussfolgerungen mitteilte. Die ersten zwei Minuten verbrachte die KI damit, die einschlägige Literatur auf dem Gebiet zu durchsuchen. Dann erklärte das Programm, dass es zunächst versuchen wollte, eine einfachere Version der Frage zu lösen. Einige Minuten später verkündete es, dass es nun bereit sei, das schwierigere Problem anzugehen. Nach fünf Minuten präsentierte o4-mini eine korrekte Lösung. »Das Modell wurde dabei richtig frech«, sagt Ono, der freiberuflich als mathematischer Berater für Epoch AI arbeitet. »Am Ende gab es aus: ›Keine Quelle nötig, denn die mysteriöse Zahl wurde von mir berechnet!‹«

Niedergeschlagen wandte sich Ono in dieser Nacht auf Signal an die anderen Teilnehmer. »Ich war nicht darauf vorbereitet, mit einem solchen LLM zu konkurrieren«, erinnert er sich. »Diese Art von Argumentation habe ich bei KI-Modellen noch nie gesehen. Es tat das, was ein Wissenschaftler tut. Das ist beängstigend.«

Obwohl es den Forschenden schließlich gelang, zehn Fragen zu finden, die dem Chatbot Schwierigkeiten bereiteten, waren sie erstaunt, wie weit die KI innerhalb eines Jahres fortgeschritten ist. Ono vergleicht das Modell mit einem fähigen Mitarbeiter. Der Mathematiker Yang Hui He vom Londoner Institute for Mathematical Sciences urteilt: »Die KI tut das, was ein sehr, sehr guter Doktorand tun würde – sogar mehr.«

Der Bot war viel schneller als ein professioneller Mathematiker und brauchte nur wenige Minuten für das, wofür ein menschlicher Experte Wochen oder Monate benötigt hätte.

Das Sparring mit o4-mini war zwar aufregend, aber die Fortschritte sind auch alarmierend. Ono und He befürchten, dass man den Ergebnissen von o4-mini zu viel Vertrauen schenken könnte. »Es gibt den Beweis durch Induktion, den Beweis durch Widerspruch und den Beweis durch Einschüchterung«, sagt He. »Wenn man etwas mit genügend Autorität sagt, bekommen die Leute einfach Angst. Ich denke, o4-mini hat den Beweis durch Einschüchterung gemeistert; es verkündet alles mit sehr viel Selbstvertrauen.«

»Die großen Sprachmodelle übertreffen bereits die Leistungen der besten Doktoranden auf der Welt«

Ken Ono, Mathematiker

Am Ende des Treffens überlegte die Gruppe, wie die Zukunft der Mathematik aussehen könnte. Die Diskussionen drehten sich um die unvermeidlichen »Tier Five«-Fragen, die selbst die besten Mathematiker auf der Welt nicht lösen können. Wenn die KI dieses Niveau erreicht, würde sich die Rolle der Menschen in dem Fach stark verändern. Sie könnten zum Beispiel dazu übergehen, einfach Fragen zu stellen und mit logisch denkenden Bots zu interagieren, um ihnen bei der Entdeckung neuer mathematischer Wahrheiten zu helfen – ähnlich wie es ein Professor mit seinen Studierenden tut. Ono geht davon aus, dass die Förderung der Kreativität in der Hochschulbildung essenziell für den Fortbestand der Mathematik künftiger Generationen sein wird.

»Ich habe meinen Kollegen immer wieder gesagt, dass es ein großer Fehler ist, davon auszugehen, dass es nie eine verallgemeinerte künstliche Intelligenz geben wird, sondern nur einen Computer«, sagt Ono. »Ich möchte nicht zur Hysterie beitragen, aber die großen Sprachmodelle übertreffen bereits die Leistungen der besten Doktoranden auf der Welt.«