

Pressemitteilung**Technische Universität Darmstadt****Silke Paradowski**

12.08.2024

<http://idw-online.de/de/news838047>Forschungsergebnisse, Wissenschaftliche Publikationen
Gesellschaft, Informationstechnik, Sprache / Literatur
überregionalTECHNISCHE
UNIVERSITÄT
DARMSTADT**Doch (noch) kein selbstständiges, komplexes Denken: Studie unter Leitung der TU zeigt Begrenztheit von ChatGPT & Co.**

Darmstadt, 12. August 2024. KI-Modelle wie ChatGPT sind laut einer neuen Studie unter führender Beteiligung der TU Darmstadt offenbar weniger selbstständig lernfähig als bisher angenommen. Es gebe keine Hinweise darauf, dass die sogenannten Large Language Models (LLMs) anfangen, ein allgemeines „intelligentes“ Verhalten zu entwickeln, das ihnen etwa ein planvolles oder intuitives Vorgehen oder komplexes Denken ermögliche, heißt es in der Untersuchung. Die Studie wird im August auf der Jahrestagung der renommierten Association for Computational Linguistics (ACL) in Bangkok vorgestellt, der größten internationalen Konferenz zur Automatischen Sprachverarbeitung.

Im Mittelpunkt der Forschung stehen unvorhergesehene und plötzliche Leistungssprünge der Sprachmodelle, die als „emergente Fähigkeiten“ bezeichnet werden. Wissenschaftler:innen hatten nach Einführung der Modelle festgestellt, dass diese mit zunehmender Größe und der Menge an Daten, mit denen sie trainiert wurden (Skalierung), leistungsfähiger wurden. So konnten die Tools mit zunehmender Skalierung eine größere Anzahl sprachbasierter Aufgaben lösen – beispielsweise gefälschte Nachrichten erkennen oder logische Schlussfolgerungen ziehen. Das weckte zum einen die Hoffnung, dass eine weitere Skalierung die Modelle noch besser machen würde. Zum anderen kam aber auch die Sorge auf, dass diese Fähigkeiten gefährlich werden könnten, da sich die LLMs quasi verselbständigen und der menschlichen Kontrolle womöglich entziehen. Als Reaktion wurden weltweit KI-Gesetze eingeführt, darunter in der Europäischen Union und in den USA.

Die Autor:innen der aktuellen Studie kommen nun allerdings zu dem Schluss, dass es für die mutmaßliche Entwicklung eines differenzierten Denkvermögens der Modelle keine Beweise gebe. Stattdessen erlangten die LLMs die oberflächliche Fertigkeit, relativ einfachen Anweisungen zu folgen, wie die Forschenden zeigten. Von dem, was Menschen können, seien die Systeme noch weit entfernt. Die Studie wurde von TU-Informatikprofessorin Iryna Gurevych und ihrem Kollegen Dr. Harish Tayyar Madabushi von der University of Bath in Großbritannien geleitet. „Unsere Ergebnisse bedeuten jedoch nicht, dass KI überhaupt keine Bedrohung darstellt“, betonte Gurevych. „Wir zeigen vielmehr, dass die angebliche Entstehung komplexer Denkfähigkeiten, die mit bestimmten Bedrohungen verbunden sind, nicht durch Beweise gestützt wird und dass wir den Lernprozess von LLMs doch gut steuern können. Daher sollte der Fokus künftiger Forschung auf weiteren Risiken liegen, die von den Modellen ausgehen, beispielsweise auf deren Potenzial, zur Generierung von Fake News genutzt zu werden.“

Und was bedeuten die Ergebnisse nun für Nutzende von KI-Systemen wie ChatGPT? „Es ist wahrscheinlich ein Fehler, sich auf ein KI-Modell zu verlassen, um komplexe Aufgaben ohne Hilfe zu interpretieren und auszuführen“, erklärt Gurevych, die an der TU Darmstadt die Arbeitsgruppe Ubiquitous Knowledge Processing (UKP) am Fachbereich Informatik leitet. „Stattdessen sollten Nutzende explizit angeben, was die Systeme tun sollen, und wenn möglich Beispiele nennen. Wichtig ist: Die Tendenz dieser Modelle, plausibel klingende, aber falsche Ergebnisse zu erzeugen – die sogenannte Konfabulation – wird wahrscheinlich weiter bestehen bleiben, auch wenn sich die Qualität der Modelle in jüngster Zeit drastisch verbessert hat.“

Über die TU Darmstadt

Die TU Darmstadt zählt zu den führenden Technischen Universitäten in Deutschland und steht für exzellente und relevante Wissenschaft. Globale Transformationen – von der Energiewende über Industrie 4.0 bis zur Künstlichen Intelligenz – gestaltet die TU Darmstadt durch herausragende Erkenntnisse und zukunftsweisende Studienangebote entscheidend mit. Ihre Spitzenforschung bündelt die TU Darmstadt in drei Feldern: Energy and Environment, Information and Intelligence, Matter and Materials. Ihre problemzentrierte Interdisziplinarität und der produktive Austausch mit Gesellschaft, Wirtschaft und Politik erzeugen Fortschritte für eine weltweit nachhaltige Entwicklung. Seit ihrer Gründung 1877 zählt die TU Darmstadt zu den am stärksten international geprägten Universitäten in Deutschland; als Europäische Technische Universität baut sie in der Allianz Unite! einen transeuropäischen Campus auf. Mit ihren Partnern der Rhein-Main-Universitäten – der Goethe-Universität Frankfurt und der Johannes Gutenberg-Universität Mainz – entwickelt sie die Metropolregion Frankfurt-Rhein-Main als global attraktiven Wissenschaftsraum weiter.
www.tu-darmstadt.de

MI-Nr. 36/2024, mih

wissenschaftliche Ansprechpartner:

Prof. Dr. Iryna Gurevych
Hochschulstraße 10, S2|02 B110, 64289 Darmstadt
Telefon: 06151/16-25290; E-Mail: iryna.gurevych@tu-darmstadt.de

Originalpublikation:

Sheng Lu, Irina Bigoulaeva, Rachneet Sachdeva, Harish Tayyar Madabushi, Iryna Gurevych: Are Emergent Abilities in Large Language Models just In-Context Learning?
<https://doi.org/10.48550/arXiv.2309.01809>
<https://arxiv.org/abs/2309.01809>