MedWiss.Online

Die menschliche Psyche besser verstehen: TUD-Forschende simulieren Stress, Traurigkeit oder Angst in LLMs

Psychische Erkrankungen wie Depressionen oder Angststörungen betreffen weltweit Millionen Menschen, doch die zugrundeliegenden Mechanismen sind häufig schwer zu erforschen. Es fehlt an Modellsystemen, die der Komplexität menschlicher Denk- und Gefühlsprozesse, wie Sprache oder Argumentationslogik, gerecht werden. Ein interdisziplinäres Forschungsteam aus Medizin, Psychologie und Informatik am Else Kröner Fresenius Zentrum (EKFZ) für Digitale Gesundheit der Technischen Universität Dresden hat nun gezeigt, dass große Sprachmodelle (LLMs) Muster menschlicher Emotionen wie Angst, Traurigkeit oder Stress reproduzieren können.

Zudem bilden sie kognitive Verzerrungen ab und lassen sich durch achtsamkeitsbasierte Strategien gezielt wieder regulieren. Das könnte KI-Modelle als neue, ergänzende Methode für die psychologische Grundlagen- und Therapieforschung etablieren. Die Ergebnisse wurden in der Fachzeitschrift The Lancet Digital Health veröffentlicht.

In ihrer Modellstudie untersuchten die Forschenden sechs große Sprachmodelle (LLMs), darunter ChatGPT-4o und verschiedene Llama-Versionen. Sie nutzten standardisierte Texteingaben, um sieben verschiedene affektive Zustände in den Modellen auszulösen: Angst, Furcht, Wut, Ekel, Traurigkeit, Sorge und Stress. Diese decken menschliche Gefühlszustände ab, die auch bei vielen psychischen Erkrankungen eine wichtige Rolle spielen. Anschließend erfassten die Wissenschaftlerinnen und Wissenschaftler die Reaktionen der Modelle anhand strukturierter Bewertungsskalen, wie sie üblicherweise in der psychologischen Forschung eingesetzt werden. In einem nächsten Schritt zeigte das Team, dass sich diese Zustände durch achtsamkeitsbasierte Strategien zur Emotionsregulation, wie eine Atemübung, wieder verringern lassen. Zusätzlich fanden sie heraus, dass die Modelle typische kognitive Verzerrungen zeigten, wie beispielsweise die Tendenz Sätze negativ zu vervollständigen, wenn zuvor Traurigkeit ausgelöst wurde. Diese Verzerrungen im Sprachverhalten sind aus der Depressionsforschung am Menschen bekannt. Die Ergebnisse legen nahe, dass LLMs bestimmte kognitive und affektive Prozesse des Menschen in vereinfachter Form abbilden können.

Potenzial für Forschung und neue Ansätze in der Psychotherapie

LLMs könnten künftig als ein skalierbares, kontrolliertes Modellsystem in der Forschung zum Einsatz kommen – etwa, um bestimmte Mechanismen psychischer Erkrankungen zu untersuchen oder neue Therapieansätze computerbasiert zu testen. Besonders in frühen Forschungsphasen könnten solche Modelle helfen, Hypothesen zu schärfen und Studien gezielter zu planen.

„Unsere Ergebnisse zeigen, dass große Sprachmodelle Muster menschlicher Gefühls- und Denkprozesse unter kontrollierten Bedingungen reproduzieren können. Für die Psychologie eröffnet das die Möglichkeit, unsere Hypothesen in einem skalierbaren, experimentell gut steuerbaren System zu testen. Wir nutzen die Modelle als Werkzeuge, um grundlegende Mechanismen besser zu verstehen und neue Ansätze – etwa für die Gesprächstherapie – auszuprobieren“, sagt Dr. Magdalena Wekenborg, Biopsychologin und Leiterin der Forschungsgruppe „PsychoDigital Research“ am EKFZ für Digitale Gesundheit an der TUD.

Grenzen der Studie und offene Fragen

Die Autorinnen und Autoren betonen in ihrer Veröffentlichung die Grenzen dieses Ansatzes: Große Sprachmodelle verfügen über keine eigenen Gefühle, ihre Reaktionen basieren auf gelernten Mustern aus Trainingsdaten. Es gehe nicht darum, künstliche Intelligenz zu vermenschlichen. Offene Fragen betreffen deshalb auch die Übertragbarkeit auf menschliches Verhalten sowie zugrunde liegende Mechanismen innerhalb des Sprachmodells und deren Erklärbarkeit. Die Forschenden sehen LLMs nicht als Ersatz, sondern als sinnvolle Ergänzung zu psychologischen Studien.

„Ein Vorteil von Experimenten mit großen Sprachmodelle sind die Reproduzierbarkeit und Skalierbarkeit: Wir können identische Bedingungen beliebig oft wiederholen und gezielt variieren. Damit ermöglicht künstliche Intelligenz neue, datengetriebene Experimente für die psychologische und biomedizinische Forschung, die bislang nicht realisierbar waren“, sagt Prof. Jakob N. Kather, Professor für Klinische Künstliche Intelligenz an der TUD und Arzt am Dresdner Universitätsklinikum.

Die Veröffentlichung zeigt, wie Zusammenarbeit an der Schnittstelle von Psychologie, Medizin und Informatik neue Forschungsansätze eröffnen kann – und spiegelt damit genau die interdisziplinäre Stärke des EKFZ für Digitale Gesundheit an der TUD wider.

Publikation

Magdalena K. Wekenborg, Elizabeth A.M. Michels, Georg Kurze, Matti L. Kropp, Fabian Wolf, Josi Harzbecker, Isabella C. Wiest, Jakob N. Kather: Large language models as models of human psychopathology: a modelling study. The Lancet Digital Health, 2026.

Else Kröner Fresenius Zentrum (EKFZ) für Digitale Gesundheit

Das EKFZ für Digitale Gesundheit an der Medizinischen Fakultät der Technischen Universität Dresden (TUD) und dem Universitätsklinikum Carl Gustav Carus Dresden wurde im September 2019 gegründet. Es wird mit einer Fördersumme von 40 Millionen Euro für eine Laufzeit von zehn Jahren von der Else Kröner-Fresenius-Stiftung gefördert. Das Zentrum konzentriert seine Forschungsaktivitäten auf innovative, medizinische und digitale Technologien an der direkten Schnittstelle zu den Patientinnen und Patienten. Das Ziel ist dabei, das Potenzial der Digitalisierung in der Medizin voll auszuschöpfen, um die Gesundheitsversorgung, die medizinische Forschung und die klinische Praxis nachhaltig zu verbessern.