

KI denkt wie wir – und könnte unser Denken besser erklären

Forschende von Helmholtz Munich haben eine Künstliche Intelligenz entwickelt, die menschliches Verhalten erstaunlich präzise simuliert. Das Sprachmodell „Centaur“ wurde auf Basis von mehr als zehn Millionen Entscheidungen aus psychologischen Experimenten trainiert - es trifft Entscheidungen auf ähnliche Weise wie echte Menschen. Damit eröffnet es neue Wege, menschliche Kognition besser zu verstehen und psychologische Theorien gezielt zu verbessern.

in der Psychologie gibt es seit langem den Traum, menschliches Denken in seiner ganzen Vielfalt zu erklären. Doch bisherige psychologische Modelle konnten entweder nachvollziehbar erklären, wie Menschen denken – oder mit hoher Genauigkeit vorhersagen, wie sie sich verhalten. Beides gleichzeitig galt bisher als kaum möglich. Das Team um Dr. Marcel Binz und Dr. Eric Schulz, beide Wissenschaftler am Institute for Human-Centered AI bei Helmholtz Munich, hat nun ein Modell entwickelt, das beides verbindet: Centaur wurde mithilfe eines eigens kuratierten Datensatzes namens Psych-101 trainiert. Er umfasst über zehn Millionen Einzelentscheidungen aus 160 Verhaltensexperimenten.

Das Besondere an Centaur: Es kann menschliches Verhalten nicht nur in bekannten Aufgaben vorhersagen, sondern auch in neuen Situationen, die es nie zuvor gesehen hat. Dabei erkennt es typische Entscheidungsstrategien, reagiert flexibel auf Kontextwechsel – und sagt selbst Reaktionszeiten erstaunlich präzise voraus. „Wir haben ein Werkzeug geschaffen, das erlaubt, menschliches Verhalten in beliebigen sprachlich beschriebenen Situationen vorherzusagen – wie ein virtuelles Versuchslabor“, sagt Marcel Binz, der auch Erstautor der Studie ist. Die Anwendungsmöglichkeiten reichen von der Analyse klassischer psychologischer Experimente bis hin zur Simulation individueller Entscheidungsprozesse in klinischen Kontexten, etwa bei Depression oder Angststörungen. Gerade für die Gesundheitsforschung eröffnet das Modell neue Perspektiven – etwa um zu verstehen, wie Menschen mit unterschiedlichen psychischen Belastungen Entscheidungen treffen. Dazu soll der Datensatz künftig auch demografische und psychologische Merkmale enthalten.

Centaur: Brücke zwischen Theorie und Vorhersage

Centaur verbindet zwei bislang getrennte Welten: verständliche Theorien und präzise Vorhersagekraft. Es kann aufdecken, wo klassische Modelle versagen – und liefert Hinweise, wie sie sich verbessern lassen. Damit eröffnet es neue Möglichkeiten in der Forschung und Anwendung, etwa in der Medizin, den Umweltwissenschaften oder den Sozialwissenschaften. „Wir stehen erst am Anfang und sehen schon jetzt enormes Potenzial“, sagt Institutsdirektor Schulz. Wichtig sei dabei auch, solche Systeme transparent und kontrollierbar zu gestalten, ergänzt Binz – etwa durch offen zugängliche, lokal gehostete Modelle, bei denen die Datenhoheit gewahrt bleibt.

Im nächsten Schritt wollen die Forschenden nun tiefer in das „Innenleben“ von Centaur blicken: Welche Rechenmuster im Modell stehen für bestimmte Entscheidungsprozesse? Lässt sich daraus ableiten, wie Menschen Informationen verarbeiten – oder wie sich Denkstrategien zwischen gesunden und erkrankten Personen unterscheiden? Die Wissenschaftler sind überzeugt: „Solche Modelle werden unser Verständnis von menschlichem Denken fundamental erweitern – wenn wir verantwortungsvoll mit ihren Möglichkeiten umgehen.“ Dass diese Forschung bei Helmholtz Munich

entsteht und nicht in den Entwicklungsabteilungen großer Tech-Konzerne, ist kein Zufall: „Wir verbinden KI-Forschung mit psychologischer Theorie – und mit einer klaren ethischen Verantwortung“, sagt Binz. „In einem öffentlichen Forschungsumfeld haben wir die Freiheit, auch grundlegende kognitive Fragen zu verfolgen, die für Unternehmen oft nicht im Fokus stehen.“

Infobox: Was ist Psych-101?

Psych-101 ist ein Datensatz, den das Team um Marcel Binz speziell für das Training des KI-Modells Centaur zusammengestellt hat. Er enthält über zehn Millionen Einzelentscheidungen von mehr als 60.000 Versuchspersonen, die an 160 psychologischen Experimenten teilgenommen haben. Die Experimente decken ein breites Spektrum menschlichen Verhaltens ab – von Risikoentscheidungen über Belohnungslernen bis hin zu moralischen Dilemmata. Die Forschenden mussten alle Daten manuell aufbereiten und in ein einheitliches Format bringen, sodass sie von einem Sprachmodell verarbeitet werden können. Damit stellt Psych-101 eine bislang einzigartige Ressource dar, um menschliches Verhalten auf Basis von Spracheingaben systematisch zu modellieren.

Über die Forschenden

Dr. Marcel Binz und Dr. Eric Schulz forschen beide am Institute for Human-Centered AI bei Helmholtz Munich.

Über Helmholtz Munich

Helmholtz Munich ist ein biomedizinisches Spitzenforschungszentrum. Seine Mission ist, bahnbrechende Lösungen für eine gesündere Gesellschaft in einer sich schnell verändernden Welt zu entwickeln. Interdisziplinäre Forschungsteams fokussieren sich auf umweltbedingte Krankheiten, insbesondere die Therapie und die Prävention von Diabetes, Adipositas, Allergien und chronischen Lungenerkrankungen. Mittels künstlicher Intelligenz und Bioengineering transferieren die Forschenden ihre Erkenntnisse schneller zu den Patient:innen. Helmholtz Munich zählt rund 2.500 Mitarbeitende und hat seinen Sitz in München/Neuherberg. Es ist Mitglied der Helmholtz-Gemeinschaft, mit mehr als 43.000 Mitarbeitenden und 18 Forschungszentren die größte Wissenschaftsorganisation in Deutschland. Mehr über Helmholtz Munich (Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt GmbH): www.helmholtz-munich.de

Originalpublikation:

Binz et al., 2025: A foundation model to predict and capture human cognition. Nature. DOI: 10.1038/s41586-025-09215-4.
<https://www.nature.com/articles/s41586-025-09215-4>