

Wie KI die Welt wahrnimmt

Untersuchung von Neurowissenschaftlern der Justus-Liebig-Universität Gießen und des Max-Planck-Instituts für Kognitions- und Neurowissenschaften

Künstliche Intelligenz (KI) wird immer wichtiger und ist bereits in vielen Aspekten unseres täglichen Lebens präsent – aber nimmt KI die Welt genauso wahr wie wir Menschen und denkt über sie nach? Um diese Frage zu beantworten, hat ein Forschungsteam der Justus-Liebig-Universität Gießen (JLU) und des Max-Planck-Instituts für Kognitions- und Neurowissenschaften untersucht, ob KI Objekte ähnlich wie Menschen erkennt, und ihre Ergebnisse in der Zeitschrift *Nature Machine Intelligence* veröffentlicht. Die Erstautoren Florian Mahner und Lukas Muttenthaler sowie Prof. Dr. Martin Hebart entwickelten einen neuen Ansatz, der es erlaubt, die Schlüsseldimensionen, auf die Menschen und KI beim Sehen von Objekten achten, eindeutig zu identifizieren und zu vergleichen.

„Diese Dimensionen repräsentieren verschiedene Eigenschaften von Objekten, die von rein visuellen Aspekten wie ‚rund‘ oder ‚weiß‘ bis hin zu eher semantischen Eigenschaften wie ‚tierbezogen‘ oder ‚feuerbezogen‘ reichen, wobei viele Dimensionen sowohl visuelle als auch semantische Elemente enthalten“, erklärt Florian Mahner. „Unsere Ergebnisse zeigen einen wichtigen Unterschied: Während sich Menschen vor allem auf Dimensionen konzentrieren, die mit der Bedeutung zusammenhängen – was ein Objekt ist und was wir darüber wissen –, verlassen sich KI-Modelle stärker auf Dimensionen, die visuelle Eigenschaften erfassen, wie etwa die Form oder Farbe des Objekts. Wir nennen dieses Phänomen ‚visuelle Bevorzugung‘ in der KI. Selbst wenn die KI Objekte genauso zu erkennen scheint wie der Mensch, wendet sie oft grundlegend andere Strategien an. Dieser Unterschied ist von Bedeutung, denn er meint, dass KI-Systeme, obwohl sie sich ähnlich wie Menschen verhalten, möglicherweise ganz anders denken und Entscheidungen treffen, was sich darauf auswirkt, wie sehr wir ihnen vertrauen können.“

Für das menschliche Verhalten nutzten die Wissenschaftler rund 5 Millionen öffentlich verfügbare Odd-One-Out-Urteile über 1.854 verschiedene Objektbilder. Einem Teilnehmer wurde zum Beispiel das Bild einer Gitarre, eines Elefanten und eines Stuhls gezeigt und er wurde gefragt, welches Objekt nicht übereinstimmt. Die Wissenschaftler nutzten dann mehrere tiefe neuronale Netze (DNNs), die Bilder analog zu menschlichen Teilnehmern erkennen können, und sammelten Ähnlichkeitsurteile für Bilder der gleichen Objekte, die für Menschen verwendet wurden. Anschließend wendeten sie denselben Algorithmus an, um die Schlüsselmerkmale dieser Bilder – von den Wissenschaftlern als „Dimensionen“ bezeichnet – zu ermitteln, die den Ausreißerentscheidungen zugrunde liegen. Dadurch, dass das neuronale Netz analog zum Menschen behandelt wurde, war eine direkte Vergleichbarkeit zwischen den beiden gewährleistet.

„Als wir uns die Dimensionen, die wir in den tiefen neuronalen Netzen entdeckten, zum ersten Mal ansahen, dachten wir, dass sie denen des Menschen sehr ähnlich sind“, erklärt

Prof. Hebart, Letztautor der Studie. „Als wir jedoch die übereinstimmenden Dimensionen zwischen Menschen und tiefen neuronalen Netzen direkt verglichen, stellten wir fest, dass das Netz diese Dimensionen nur annähernd erfasste. Bei einer tierbezogenen Dimension wurden viele Bilder von Tieren nicht einbezogen, und ebenso wurden viele Bilder einbezogen, die gar keine Tiere waren. Das ist etwas, das wir mit Standardtechniken übersehen hätten.“

Die Wissenschaftler hoffen, dass zukünftige Forschungen ähnliche Ansätze verwenden werden, die Menschen und KI direkt miteinander vergleichen, um besser zu verstehen, wie KI die Welt wahrnimmt. „Unsere Forschung bietet eine klare und interpretierbare Methode zur Untersuchung dieser Unterschiede, die uns hilft, besser zu verstehen, wie KI im Vergleich zum Menschen Informationen verarbeitet“, sagt Martin Hebart. „Dieses Wissen kann uns nicht nur helfen, die KI-Technologie zu verbessern, sondern liefert auch wertvolle Einblicke in die menschliche Kognition.“

Publikation

Mahner, F.P., Muttenthaler, L., Güçlü, U., Hebart, M.: Dimensions underlying the representational alignment of deep neural networks with humans. *Nat Mach Intell* (2025). <https://www.nature.com/articles/s42256-025-01041-7>

Kontakt

Prof. Dr. Martin Hebart

Computational Cognitive Neuroscience and Quantitative Psychiatry

E-Mail: martin.hebart@psychiat.med.uni-giessen.de

Die 1607 gegründete **Justus-Liebig-Universität Gießen (JLU)** ist eine traditionsreiche Forschungsuniversität, die rund 25.000 Studierende anzieht. Neben einem breiten Lehrangebot – von den klassischen Naturwissenschaften über Rechts- und Wirtschaftswissenschaften, Gesellschafts- und Erziehungswissenschaften bis hin zu Sprach- und Kulturwissenschaften – bietet sie ein lebenswissenschaftliches Fächerspektrum, das nicht nur in Hessen einmalig ist: Human- und Veterinärmedizin, Agrar-, Umwelt- und Ernährungswissenschaften sowie Lebensmittelchemie. Unter den großen Persönlichkeiten, die an der JLU geforscht und gelehrt haben, befindet sich eine Reihe von Nobelpreisträgern, unter anderem Wilhelm Conrad Röntgen (Nobelpreis für Physik 1901) und Wangari Maathai (Friedensnobelpreis 2004). Seit dem Jahr 2006 wird die Forschung an der JLU kontinuierlich in der Exzellenzinitiative bzw. der Exzellenzstrategie von Bund und Ländern gefördert.