

## Pressemitteilung

Albert-Ludwigs-Universität Freiburg im Breisgau

Rimma Gerenstein

18.07.2024

<http://idw-online.de/de/news837175>

universität freiburg

Forschungsergebnisse  
Informationstechnik, Psychologie  
überregional

## Interne Weltmodelle von Menschen, Tieren und KI untersuchen

**- Freiburger Forschende erarbeiten neue formale Beschreibung interner Weltmodelle und vereinfachen damit interdisziplinäre Forschung - Interne Weltmodelle helfen, aus bisherigen Erfahrungen Vorhersagen über neue Situationen zu treffen und sich so zurechtzufinden - Die neue Formalisierung hilft, Weltmodelle von Menschen, Tieren und KI zu vergleichen und Defizite zu beheben**

Ein Team aus Wissenschaftler\*innen um Prof. Dr. Ilka Diester, Professorin für Optophysiology und Sprecherin des Forschungszentrums BrainLinks-BrainTools an der Universität Freiburg, hat eine formale Beschreibung interner Weltmodelle entwickelt und diese in der Zeitschrift „Neuron“ veröffentlicht. Die Formalisierung hilft Wissenschaftler\*innen, die Entstehung und Funktionsweise interner Weltmodelle besser zu verstehen. Sie ermöglicht, Weltmodelle von Menschen, Tieren und Künstlicher Intelligenz (KI) systematisch miteinander zu vergleichen. So wird beispielsweise klarer, wo KI im Vergleich zu menschlicher Intelligenz noch Defizite aufweist und wie sie zukünftig weiterentwickelt werden könnte. An der interdisziplinären Publikation waren elf Freiburger Forschende aus vier Fakultäten beteiligt.

Interne Weltmodelle: aus Erfahrungen Vorhersagen treffen

Aus täglichen Erfahrungen abstrahieren Menschen und Tiere allgemeine Gesetzmäßigkeiten. Sie entwickeln interne Weltmodelle, die ihnen helfen, sich auch in unbekanntem Zusammenhängen zurechtzufinden. Denn auf Basis der abstrahierten Modelle können sie in neuen Situationen Vorhersagen treffen und sich dementsprechend verhalten. Beispielsweise hilft es bei der Orientierung in einer fremden Stadt, vergleichbare Städte zu kennen, in denen es etwa auch ein Stadtzentrum, Fußgängerzonen und öffentlichen Nahverkehr gibt. Auch in sozialen Kontexten wie beim Abendessen im Restaurant helfen vergleichbare Erfahrungen, um sich angemessen zu verhalten.

Weltmodelle werden mit Hilfe des neuen Formalismus greifbarer

Um interne Weltmodelle über Spezies hinweg zu formalisieren, unterscheiden die Forschenden in ihrer aktuellen Publikation zwischen drei abstrakten Räumen, die ineinandergreifen: dem Aufgabenraum, dem neuronalen Raum und dem konzeptionellen Raum. Der Aufgabenraum umfasst all das, was ein Individuum erlebt. Der neuronale Raum beschreibt die verschiedenen messbaren Zustände des Hirns, von der molekularen Ebene über die Aktivität einzelner Neurone bis hin zur Aktivität ganzer Hirnareale. Letztere wird beispielsweise mit Hilfe eines funktionellen Magnetresonanztomographen (fMRT) sichtbar oder mittels Techniken wie hochdichten Elektroden oder Kalziumbildgebung gemessen. Das Äquivalent des neuronalen Raums bei KI ist die Aktivität der Knoten innerhalb des entsprechenden künstlichen neuronalen Netzes. Der konzeptionelle Raum besteht aus Paaren von Zuständen des Aufgaben- und des neuronalen Raums. Diese Paare stellen somit den Status eines Individuums dar, der interne Prozesse mit externen Einflüssen verknüpft. Der aktuelle Status verändert sich ständig, indem er mit einer bestimmten Wahrscheinlichkeit in den nächsten übergeht. Diese Kombinationen aus den Erfahrungen eines Individuums einerseits und der dazugehörigen Hirnaktivität andererseits sowie die dynamischen Übergänge machen die individuellen internen Weltmodelle wissenschaftlich greifbar.

## Defizite interner Weltmodelle beheben

Mithilfe des Formalismus können Wissenschaftler\*innen nun über Disziplingrenzen hinweg interne Weltmodelle untersuchen und sich darüber austauschen, wie sie entstehen und sich weiterentwickeln. Erkenntnisse aus der Forschung zu Menschen und Tieren sollen beispielsweise helfen, KI zu verbessern. So sind heutige KI-Systeme noch nicht in der Lage, die Plausibilität ihrer Vorhersagen zu prüfen. Auch Large Language Models wie ChatGPT funktionieren bisher nur als Mustererkennungsmaschinen ohne die Fähigkeit, wirklich zu planen. Planung ist jedoch wichtig, um in unbekanntem Situationen Strategien durchzuspielen und zu korrigieren, bevor sie umgesetzt werden und dabei gegebenenfalls Schaden anrichten. Forscher\*innen vermuten außerdem, dass Defizite in internen Weltmodellen die Ursache einiger psychischer Erkrankungen wie Depressionen oder Schizophrenie sein könnten. Hier könnte ein tieferes Verständnis von Weltmodellen helfen, Medikation und Therapie gezielter einzusetzen.

Original-Publikation: Diester et al., Internal world models in humans, animals, and AI, Neuron (2024), <https://doi.org/10.1016/j.neuron.2024.06.019>

Prof. Dr. Ilka Diester ist Professorin für Optophysiology an der Universität Freiburg und leitet an der Fakultät für Biologie das Optophysiology Lab. Zudem ist Prof. Diester Sprecherin des Forschungszentrums BrainLinks-BrainTools//IMBIT sowie des Forschungsfelds Gehirn und Intelligenz der Universität Freiburg. Die meisten Ko-Autor\*innen der Studie forschen ebenfalls an der Universität Freiburg: Prof. Dr. Marlene Bartos ist Professorin für Zelluläre und Systemische Neurophysiologie und Sprecherin des SFBs IN-CODE; Prof. Dr. Joschka Bödecker ist Professor für Neurorobotik; Dr. Adam Kortylewski ist Emmy Noether-Gruppenleiter am Institut für Informatik; Prof. Dr. Christian Leibold ist Professor für Theoretische Systemneurowissenschaften; Prof. Dr. Johannes Letzkus ist Heisenberg-Professor am Institut für Physiologie; Prof. Dr. Monika Schönauer ist Juniorprofessorin für Neuropsychologie; Prof. Dr. Andrew Straw ist Professor für Neuro- & Verhaltensbiologie; Prof. Dr. Abhinav Valada ist Professor für Autonome Intelligente Systeme; Prof. Dr. Andreas Vlachos ist Professor für Neuroanatomie; Prof. Dr. Thomas Brox ist Professor für Mustererkennung und Bildverarbeitung. Die Professoren Bödecker, Valada und Brox sind Teil des Europäischen Netzwerks für Maschinelles Lernen und Intelligente Systeme ELLIS. Dr. Matthew Nour forscht an der University of Oxford im Department of Psychiatry.

Die Publikation ist Ergebnis eines Workshops der BrainWorlds-Initiative am Forschungszentrum BrainLinks-BrainTools//IMBIT an der Universität Freiburg.

Kontakt:  
Hochschul- und Wissenschaftskommunikation  
Universität Freiburg  
Tel.: (+49) 0761/203 4302  
E-Mail: [kommunikation@zv.uni-freiburg.de](mailto:kommunikation@zv.uni-freiburg.de)

wissenschaftliche Ansprechpartner:  
Prof. Dr. Ilka Diester  
Albert-Ludwigs-Universität Freiburg  
Fakultät für Biologie  
Optophysiology  
Georges-Köhler-Allee 201  
D-79100 Freiburg

Tel +49 (0)761 203-8440  
E-Mail: [ilka.diester@biologie.uni-freiburg.de](mailto:ilka.diester@biologie.uni-freiburg.de)

Originalpublikation:  
<https://doi.org/10.1016/j.neuron.2024.06.019>

URL zur Pressemitteilung: <https://www.brainlinks-braintools.uni-freiburg.de/de/>

