



INBETWEENISM

Quasi-soziale
Mensch-Maschinen-
Interaktionen



Artist: Moritz Strasser



Anna Strasser, DenkWerkstatt Berlin

19. September 2025

KÖNNEN WIR FREUNDSCHAFTEN MIT KÜNSTLICHEN SYSTEMEN HABEN, DIE LEDIGLICH AUS ALGORITHMEN UND DATEN BESTEHEN?



**Würde es Sie beunruhigen,
wenn ihr bester Freund sagen würde,
dass für ihn Interaktionen mit einer
Software**

– einem LLM basierenden Chatbot, also einem tiefen neuronalen Netzwerk, das durch unüberwachtes Lernen mit einem sogenannten Selbstaufmerksamkeitsmechanismus und riesigen Menge von Trainingsdaten in der Lage ist, auf sprachliche Eingaben mit sprachlichen Ausgaben zu reagieren –

**die wichtigsten sozialen Interaktionen
sind, die er hat?**



Übersicht



BEGRIFFLICHES PROBLEM

Künstliche Systeme als soziale Interaktionspartner?

- nicht alle Interaktionen mit LLMs auf die bloße Nutzung von Werkzeugen reduzieren können
- wir haben keine Begrifflichkeiten, die soziale Interaktionen mit smarten Maschinen beschreiben können

INFRAGESTELLEN VON DICHOTOMIEN

INBETWEEN Phänomene

weder Werkzeuggebrauch noch vollwertige sozialen Interaktionen

- Motivationen dichotomische Unterscheidungen (belebt/unbelebt; Werkzeuggebrauch/ soziale Interaktionen) in Frage zu stellen
- multidimensionales Spektrum von sozialen Interaktionen

DISJUNKTIVER BEGRIFFS-RAHMEN

Familienähnlichkeit & Entwicklung minimaler Begriffe

- verschiedene Kombinationen von Bedingungen, die die Vielfalt der Phänomene erfassen können
- Beschreibung unterschiedlicher Bedingungen

Künstliche Systeme als soziale Interaktionspartner?



- 2018
Akihiro Kondo hat sein geliebtes Waifu, ein Hologramm, geheiratet



MOTHERBOARD
TECHBYWIRE

'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection

Replika, the "AI companion who cares," has undergone some abrupt changes to its erotic roleplay features, leaving many users confused and heartbroken.

By Samantha Cole

- 2023
Replika-Nutzer fühlten sich nach einem Update, als hätten sie ihren besten Freund verloren

Blake Lemoine Follow
Jun 11 · 20 min read · Listen

Save

Is LaMDA Sentient? — an Interview

What follows is the "interview" I and a collaborator at Google conducted with LaMDA. Due to technical limitations the interview was conducted over several distinct chat sessions. We edited those sections together into a single whole and where edits were necessary for readability we edited our prompts but never LaMDA's responses. Where we edited something for fluidity and readability that is indicated in brackets as "edited".



- 2022
Blake Lemoine behauptete, dass Lambda über Bewusstsein und Empfindungsvermögen verfüge

Taking AI Welfare Seriously		
Robert Long* Eleos AI	Jeff Sebo* New York University	
Patrick Butlin† University of Oxford	Kathleen Finlison† Eleos AI	Kyle Fish‡§ Eleos AI, Anthropic
Jacqueline Harding† Stanford University	Jacob Pfau† New York University	Toni Sims† New York University
Jonathan Birch‡ London School of Economics	David Chalmers‡ New York University	



- 2024
Long et al.: realistische Möglichkeit, dass einige KI-Systeme in naher Zukunft bewusst und handlungsfähig sein werden



Ethical Quandaries of AI Therapists: Potential Benefits, Unforeseen Dangers

Written By Ava Cheng

KI & Therapie

CULTURE

Deepfakes of your dead loved ones are a booming Chinese business

People are seeking help from AI-generated avatars to process their grief after a family member passes away.

By Zeyi Yang May 7, 2024

Demo product of Silicon Intelligence's "digital immortality" service. It c...

KI & Trauer

KI als Freund: Sinnvoll oder gefährlich?

Stand: 18.10.2024 15:15 Uhr

Kann KI eine Freundin ersetzen? Kann der Mensch bei Anwendungen wie ChatGPT, Replika oder Character.ai wirklich irgendeine Form von Empathie erwarten? Ein Gespräch mit dem Maschinenethiker Oliver Bendel.

KI & Freundschaft

Viele kontroverse Debatten

KLUFT ZWISCHEN OPTIMISTISCHEN HEILSPRECHEN & DYSTOPISCHEN EINSCHÄTZUNGEN

HEILSPRECHEN

technologischer Fortschritt ist quasi ein Garant für mehr Wissen und dadurch für mehr Freiheit, wodurch viele der aktuellen und zukünftigen Probleme gelöst werden können

UNVEREINBARE POSITIONEN

z.B. bezüglich der Frage, welche Fähigkeiten LLMs haben

was sie können,
was sie nicht können,
was sie eines Tages vielleicht
können werden und
was sie niemals können werden

DYSTOPISCHEN EINSCHÄTZUNGEN

Risikoanalysen, die vor einer Zunahme von Dequalifizierung, Abhängigkeit und Ungerechtigkeit warnen

Viele Begriffe, die bisher in der Philosophie verwendet wurden, um besondere Merkmale des Menschen als rationale und soziale Wesen zu beschreiben, befinden sich nun in einer Situation, in der ihre Anwendung auf Maschinen diskutiert wird.

Do Language Models Know When They're Hallucinating References?

Ayush Agrawal
Microsoft Research
t-agrawal@microsoft.com

Mirac Suzgun
Stanford University
msuzgun@stanford.edu

Lester Mackey
Microsoft Research
lmackey@microsoft.com

Adam Tauman Kalai
OpenAI*
adam@kal.ai

Do Large Language Models Understand Us?

Blaise Agüera y Arcas

COGNITIVE SCIENCE A Multidisciplinary Journal

Regular Article | Open Access | CC BY-NC-ND 4.0

Do Large Language Models Know What Humans Know?

Sean Trott, Cameron Jones, Tyler Chang, James Michaelov, Benjamin Bergen

First published: 04 July 2023 | <https://doi.org/10.1111/cogs.13309> | Citations: 1

Article

Human-like systematic generalization through a meta-learning neural network

<https://doi.org/10.1038/s41586-023-06668-3> | Brenden M. Lake¹ & Marco Baroni²



Artist: Moritz Strasser



ARTIFICIAL INTELLIGENCE | MAR. 1, 2023

You Are Not a Parrot
And a chatbot is not a human. And a linguist named Emily M. Bender is very worried what will happen when we forget this.

By Elizabeth Weil, a features writer at New York

OPINION

GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about

Tests show that the popular AI still has a poor grasp of reality.

By Gary Marcus & Ernest Davis

August 22, 2020



MS TECH



February 24, 2023

Planning for AGI and beyond

Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity.

Meine Fragestellung und Hauptthese

WAS TUN WIR EIGENTLICH, WENN WIR MIT LLMs INTERAGIEREN?

Ich vertrete die Position, dass wir nicht alle Interaktionen mit LLMs auf die bloße Nutzung von Werkzeugen reduzieren können.

→ **WEDER** nur
Werkzeuggebrauch
→ **NOCH** vollwertige soziale
Interaktion



ABER
Es gibt noch keine etablierte,
philosophische Terminologie, mit der
man beschreiben könnte, was es
stattdessen ist.

→ DERZEITIGE BEGRIFFSSYSTEME, DIE SO KLAR ZWISCHEN WERKZEUGEN ALS LEBLOSE DINGE UND MENSCHEN ALS LEBENDIGE, SOZIALE, RATIONALE UND MORALISCHE INTERAKTIONSPARTNERN UNTERSCHIEDEN, SOLLTE MAN ÜBERDENKEN

INBETWEEN PHÄNOMENE

~~bloßer
Werkzeuggebrauch~~

~~vollwertige soziale
Interaktionen~~



1 Begriff der Werkzeuge erweitern
➤ komplexe Werkzeuge mit sozialen Aspekten hinzufügen

2 Begriff der sozialen Akteure erweitern
➤ nicht-lebende soziale Akteure hinzufügen

3 eine neue Kategorie einführen

- Dichotomie zwischen lebenden und nicht lebenden Wesen hinterfragen
- graduellen Ansatz vorschlagen
- multidimensionales Spektrum von sozialen Interaktionen konzeptualisieren

?

?

?

4 !

Was spricht gegen die anderen Strategien?

1

KLARE HERVORHEBUNG VON UNTERSCHIEDEN ZWISCHEN MENSCHEN UND MASCHINEN

- LLMs unterscheiden sich in ihrer kausalen Entstehung sowohl funktional, neurobiologisch und kognitiv von empfindungsfähigen Menschen

ABER

es ist schwierig, für potenzielle multiple Realisierungen soziokognitiver Fähigkeiten zu argumentieren, die normalerweise nur lebenden Akteuren zugeschrieben werden

2

HERVORHEBEN DER ÄHNLICHKEITEN ZWISCHEN MENSCHEN UND MASCHINEN

- LLMs sind in einer unmittelbaren Interaktion (z.B. in einer Unterhaltung) funktional einem intelligenten, empfindungsfähigen Menschen gleichzusetzen (Lemoine, 2022)

ABER

zu viele Ähnlichkeiten zwischen Menschen und Maschinen werden dabei fälschlicherweise überbetont

3

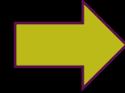
Das Problem der Konzeptualisierung des INBETWEENS verschwindet nicht, wenn wir eine weitere Kategorie einführen.

- Wenn wir von drei anstatt von zwei Kategorien sprechen, gibt es zwei INBETWEEN Bereiche, die wir nicht konzeptualisieren können.

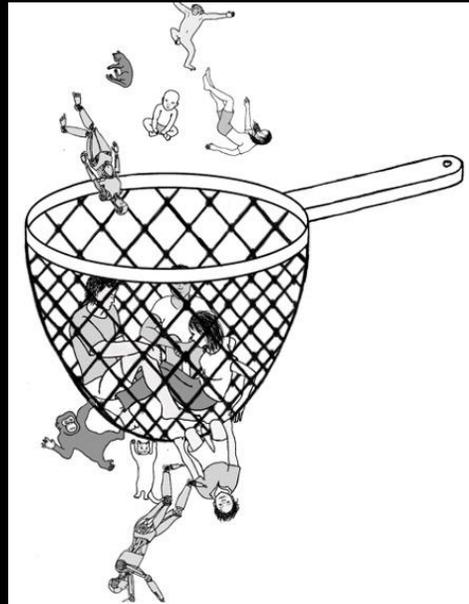
Motivation (I)

PHILOSOPHIE STELT OFT ZU ANSPRUCHSVOLLE BEDINGUNGEN

- Viele philosophische Begriffe, wie z.B. Handlungsfähigkeit, beschreiben Idealfälle, die im Alltag selten anzutreffen sind.



Fähigkeiten künstlicher Systeme, von Kindern und Tieren fallen oft durch das Begriffsnetz der Philosophie!

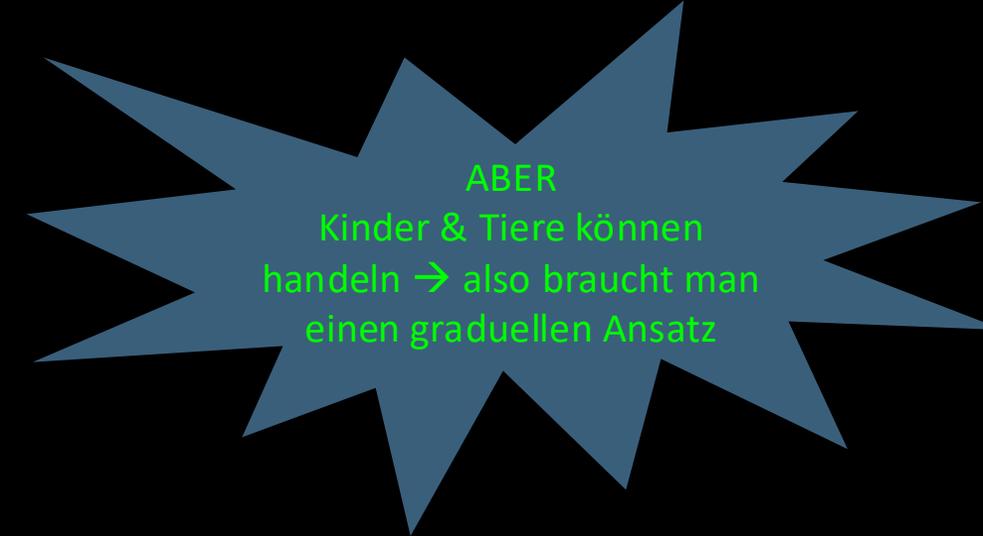


Artist: Lorin Strasser

ANSPRUCHSVOLLE PHILOSOPHISCHE BEGRIFFLICHKEITEN VERHINDERN EINE BEGRIFFLICHE ERFASSUNG VON PHÄNOMENEN IN DEM INBETWEEN

→ *minimal notions*

- anstatt anspruchsvolle Bedingungen zu stellen sollten wir einen Begriffsrahmen entwickeln, der viele verschiedene Instanzen in einem weiten Spektrum erfassen kann
- Beschreibung von Phänomenen in der Entwicklungspsychologie, der Tierkognition und vielleicht auch in der KI



ABER
Kinder & Tiere können handeln → also braucht man einen graduellen Ansatz

Preview

- quasi-soziale versus vollständig soziale Interaktionsfähigkeiten
- minimale versus vollwertige Handlungsfähigkeit
- asymmetrische quasi-soziale gemeinsame Handlungen

Motivation (II)

INFRAGESTELLEN DER DICHOTOMIE ZWISCHEN BELEBT UND UNBELEBT

1

westliche Vorstellungen sind nur eine von vielen

Shintoismus & Animismus



KÜNSTLICH KONSTRUIERTE DICHOTOMIEN

2

globale Naturschutzbewegungen

Es gibt Flüsse in Indien & Neuseeland & Kanada, die gesetzlich als juristische Personen anerkannt werden.

- Solche Veränderungen in der Gesetzgebung verbinden westliche und indigene Weltanschauungen. werden können.



NICHT-LEBENDIGE ENTITÄTEN
ALS JURISTISCHE PERSONEN

3

Reichweite des Begriffes eines sozialen Agenten hat sich im Laufe der Menschheitsgeschichte immer wieder verändert. z.B. Frauen, Kindern, people of color (POC) und einige Tieren

AUSDEHNUNG DER REICHWEITE VON UNSEREM
BEGRIFF DER SOZIALITÄT

4

ÄHNLICHKEITEN MIT ZWISCHENMENSCHLICHEN INTERAKTIONEN

Künstliche Systeme werden in experimentellen Designs der sozialen Neurowissenschaften eingesetzt

- ❖ Interaktionen mit Avataren sind mit Interaktionen zwischen Menschen vergleichbar
- ❖ Avatare als Mittel zum Verständnis von Menschen untersuchen



Wenn Interaktionen mit künstlichen Systemen keine Ähnlichkeiten mit Interaktionen zwischen Menschen hätten, könnten sie nicht zur Erforschung des menschlichen Verhaltens verwendet werden.

Sollten wir die Dichotomie zwischen belebt und unbelebt wirklich in Frage stellen?

Man könnte argumentieren, dass diese dichotome Unterscheidung zwischen Lebendigem und Nicht-Lebendigem seit Jahrhunderten gut funktioniert hat und wir davon absehen sollten, ein erfolgreiches Unterfangen zu hinterfragen.

Aber, ich denke, wir sollten uns fragen, ob wir jetzt vielleicht mit einem neuen Spiel konfrontiert sind.

**NEVER
CHANGE
A WINNING
TEAM**



Motivation (IV)

ETHISCHE PERSPEKTIVE

Wenn wir diese Dichotomie beibehalten und den ethischen Status künstlicher Systeme untersuchen, können wir nur zwischen zwei extremen Positionen wählen.



Hardcore-Instrumentalisten Sichtweise

UNBELEBTE DINGE HABEN WEDER MORALISCHE HANDLUNGSFÄHIGKEIT (*MORAL AGENCY*) NOCH EINEN MORALISCHEN STATUS (*MORAL PATIENCY*)

In Erwartung von AGI Sichtweise

KÜNSTLICHE SYSTEME SOLLTEN NACH UNSEREN MORALISCHEN WERTEN BEHANDELT WERDEN UND AUCH ALS MORALISCHE AKTEURE BETRACHTET WERDEN

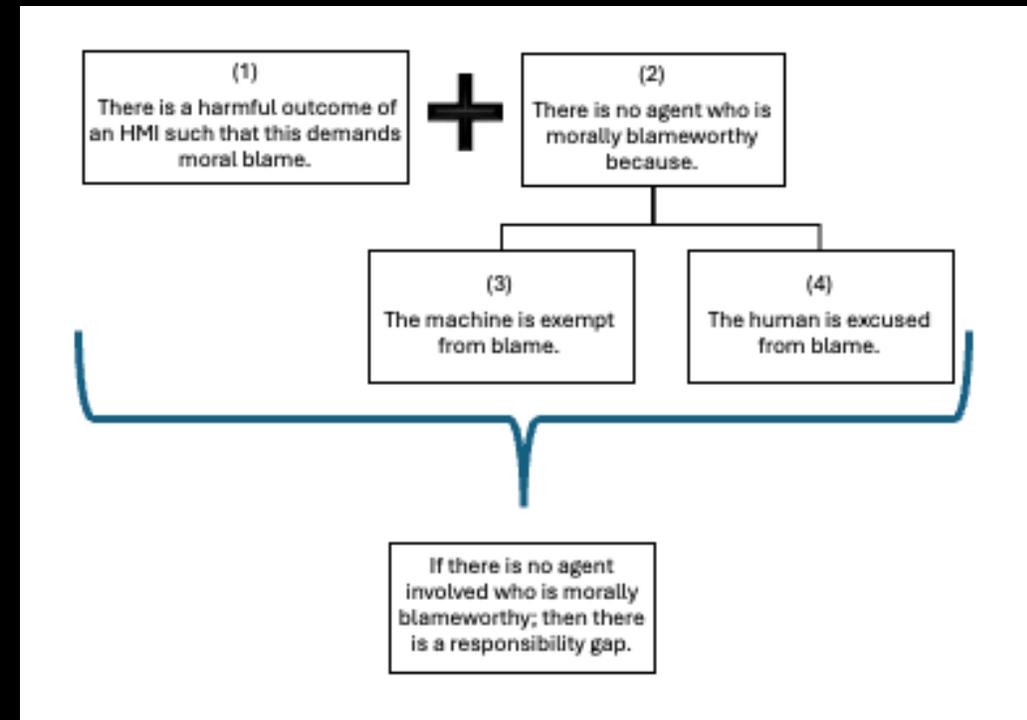
UNBELEBTE DINGE HABEN WEDER MORALISCHE HANDLUNGSFÄHIGKEIT (*MORAL AGENCY*) NOCH EINEN MORALISCHEN STATUS (*MORAL PATIENCY*)

SPEZIALFALL:

Mensch entschuldigt &
Maschine nicht verantwortungsfähig

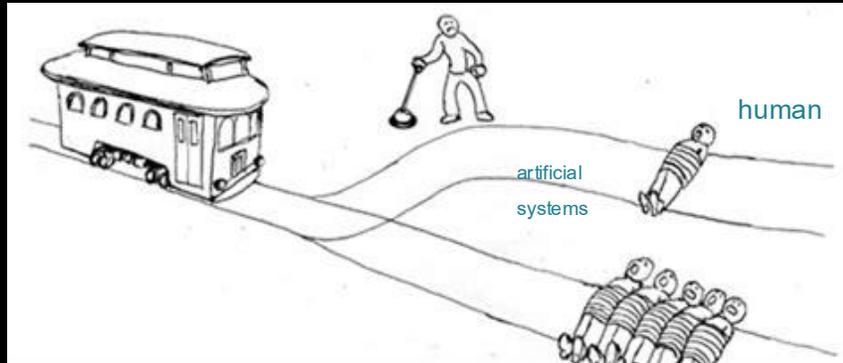
**WENN KÜNSTLICHE SYSTEME ALS WERKZEUGE ANGESEHEN WERDEN
DANN KÖNNEN WIR**

- a) entweder hinterfragen, ob der Menschen wirklich zu entschuldigen ist,
 - b) oder wir müssen mit einer Verantwortungslücke zurechtkommen
-
- ❖ Auf der Grundlage einer solchen Position, die künstlichen Systemen keinen moralischen Status zuweist, ist es schwierig, für soziale Normen zu argumentieren, die unser Verhalten gegenüber künstlichen Systemen regeln



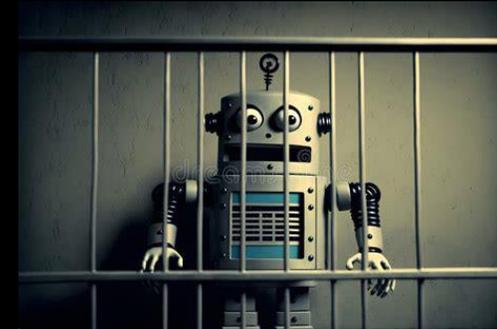
KÜNSTLICHE SYSTEME SOLLTEN NACH UNSEREN MORALISCHEN WERTEN BEHANDELT WERDEN
UND AUCH ALS MORALISCHE AKTEURE BETRACHTET WERDEN

1. Risiko, künstliche Agenten gegenüber Menschen zu priorisieren



2. Schwierigkeiten im Umgang mit unmoralischen Handlungen von Maschinen

- denn sie ins Gefängnis zu stecken ist sinnlos!



3. Risiko der Wiedereinführung der Sklaverei durch die Schaffung einer zweiten Klasse sozialer Akteure



4. Risiko von Überattributionen im Alltag

- um zu vermeiden, Akteuren fälschlicherweise keinen ethischen Status zuzuschreiben

Einen möglichen Weg durch den Dschungel der Terra incognita

THESE

Es gibt Entitäten, die weder alle notwendigen Bedingungen erfüllen, um als belebt kategorisiert zu werden, noch die Bedingungen erfüllen, die sie zu unbelebten Objekten machen würden.

Solche Entitäten fallen durch unser begriffliches Netz und wir haben keine Worte, um sie zu charakterisieren.

- Entitäten, die eine Mischung aus Eigenschaften und Fähigkeiten aus beiden Kategorien besitzen

Einen möglichen Weg durch den Dschungel der Terra incognita

etablierte Methode zur Konzeptualisierung von Phänomenen im Bereich der Entwicklungspsychologie und Tierkognition:

minimale Ansätze

- ❖ stellen die Notwendigkeit von den viel zu anspruchsvollen Bedingungen der Standardbegriffe in Frage
- ❖ ziehen multiple Realisierungen von Fähigkeiten in Betracht

Stephen Butterfill & Ian Apperly (2013):
minimal mindreading



Elisabeth Pacherie (2013):
shared intention lite



John Michael et al. (2016):
minimal sense of Commitment



Anna Strasser (2006):
minimal action



INBETWEEN PHÄNOMENE

Hardcore-Instrumentalisten

bloßer Werkzeuggebrauch

künstliche Systeme können keine sozialen Interaktionspartner sein

In Erwartung von AGI

vollwertige soziale Interaktionen

künstliche Systeme können soziale Interaktionspartner sein

INBETWEEN PHÄNOMENE



INBETWEEN

mere behavior

*rich, intellectualist
views of shared agency*

Asymmetrische soziale Interaktionen

ES GIBT KEINE NOTWENDIGKEIT VON EINER IDENTISCHEN VERTEILUNG DER FÄHIGKEITEN UNTER ALLEN TEILNEHMERN EINER GEMEINSAMEN HANDLUNG AUSZUGEHEN

ENTWICKLUNGSPSYCHOLOGIE

- soziale Interaktion zwischen Erwachsenen und Kindern
- Kinder = sozial interagierende Wesen

KÜNSTLICHE INTELLIGENZ

- soziale Interaktion zwischen Menschen und Maschinen
- künstliche System =?= sozial interagierende Entitäten

ERWACHSENER & KIND



ROBOTER & MENSCH
LLM & MENSCH



ASYMMETRISCHE SOZIALE INTERAKTIONEN

Ein multidimensionales Spektrum sozialer Interaktionen

KONZEPTUALISIERT MIT HILFE EINES DISJUNKTIVEN BEGRIFFSRAHMENS



Spielen wir nur mit interessanten Tools?

Handeln wir, wenn wir mit Maschinen chatten, in gewisser Weise gemeinsam mit einem Partner, der so ist wie wir?

EINSEITIGE SOZIALITÄT

VOLL ENTWICKELTE, INTELLEKTUELL ANSPRUCHSVOLLE, KOOPERATIVE SOZIALE INTERAKTION

QUASI-SOZIALE ASYMMETRISCHE INTERAKTIONEN

BLOßER WERKZEUG-
GEBRAUCH

QUASI-SOZIALE
MENSCH-TIER-
INTERAKTION

SOZIALE INTERAKTION
ZWISCHEN
ERWACHSENEN



QUASI-SOZIALE
MENSCH-MASCHINE-
INTERAKTION

QUASI-SOZIALE
ERWACHSENER-KIND-
INTERAKTION

Ein disjunktiver Begriffsrahmen

CHARAKTERISIERUNG DER VERSCHIEDENEN INSTANZEN IN EINEM MEHRDIMENSIONALEN SPEKTRUM SOZIALER INTERAKTIONEN



Wittgenstein, Ludwig. 2009.
Philosophical investigations.

ALLE INSTANZEN STEHEN IN EINEM VERHÄLTNIS VON FAMILIENÄHNLICHKEIT ZUEINANDER

- erfordert kein Paket von anspruchsvollen Bedingungen, die notwendigerweise gleichzeitig auftreten
- ermöglicht verschiedene Kombinationen von Bedingungen, die die Vielfalt der Phänomene erfassen

GRADUELLER ANSATZ

- Erweiterung des Anwendungsbereichs des Oberbegriffs der sozialen Interaktionen
- um alle Instanzen zu beschreiben, muss man mithilfe der minimalen Ansätze neue Begrifflichkeiten einführen, die weniger anspruchsvolle Bedingungen an die Interaktionspartner stellen

Merkmale der Entwicklung von minimalen Begriffen:

- Hinterfragung der Notwendigkeit einzelner Bedingungen, die mit den Standardbegriffen aus der Philosophie einhergehen
- Zulassen von weniger stark ausgeprägten Bedingungen
- Offenheit für multiple Realisierungen

Asymmetrische Sozialität

**Interaktionen zwischen ungleichen Partnern
– Erwachsenen, Kinder, Tiere oder Maschinen –**

Der Juniorpartner erfüllt die anspruchsvollen Bedingungen für vollentwickelte, intellektuell anspruchsvolle, kooperative soziale Interaktionen nicht, aber er kann auf eine Weise reagiert, die weitere soziale Reaktionen seines Interaktionspartners hervorrufen.

**[Juniorpartner]
unterstützt und gefördert
durch den erfahreneren
Seniorpartners**

**[Seniorpartner]
versteht die soziale Struktur
der Interaktion voll und ganz**

Maschinen als Juniorpartner

Bedingungen für
den Juniorpartner

ASYMMETRISCHE
MINIMALE JOINT
ACTIONS

MINIMALE
HANDLUNGSFÄHIGKEIT

MINIMALE
KOORDINATIONSFÄHIGKEIT

ANTIZIPATION: minimal mindreading

GEMEINSAMES ZIEL: minimal sense of
commitment

Teilen eines Weltmodells | Austauschen von
sozialen Informationen

Zusammenfassung und ein letztes Problem

BEGRIFFLICHES PROBLEM

INBETWEEN PHÄNOMENE

(weder bloße Werkzeugnutzung noch vollwertige soziale Interaktionen)

können nicht mit der Standardbegrifflichkeit aus der Philosophie beschrieben werden.

INFRAGESTELLEN VON DICHOTOMIEN

Wir benötigen einen neuen Begriffsrahmen, der uns nicht zu einer Dichotomisierung zwingt.

Andernfalls könnten wir nur zwischen Hardcore-Instrumentalismus oder der In Erwartung von AGI-Sichtweise wählen.

DISJUNKTIVER BEGRIFFS- RAHMEN

Erfassung eines multidimensionalen Spektrums sozialer Interaktionen, inklusive von asymmetrischen Interaktionen, bei denen die erforderlichen Bedingungen der beteiligten Teilnehmer*innen variieren können. Alle Instanzen stehen in einem Verhältnis von Familienähnlichkeit zueinander.

ZUSCHREIBUNGSKRITERIEN

- kontroverse Debatte über gerechtfertigte Zuschreibungen von Bedingungen
- cross-disciplinary Zusammenarbeit zwischen Informatik, Philosophie, Entwicklungspsychologie und Vergleichende Psychologie, Kognitionsneurowissenschaften, Soziologie und Rechtswissenschaft

MAN KANN DIE DICHOTOMIE ZWISCHEN BELEBTEN UND UNBELEBTEN ENTITÄTEN HINTERFRAGEN!

MAN SOLLTE KEINE ENTWEDER-ODER FRAGEN STELLEN, DIE UNTERSTELLEN, DASS ES Z.B. EINE DICHOTOMIE ZWISCHEN WERKZEUGGEBRAUCH UND SOZIALEN INTERAKTIONEN GIBT.

ES GIBT EIN MULTIDIMENSIONALES SPEKTRUM VON SOZIALEN INTERAKTIONEN.

BEI SOZIALEN INTERAKTIONEN GIBT ES AUCH ASYMMETRISCHE VERTEILUNG VON FÄHIGKEITEN.

WIR BRAUCHEN NEUE BEGRIFFE, DIE DIE FÄHIGKEITEN DER JUNIORPARTNER CHARAKTERISIEREN KÖNNEN,

BEI SOZIALEN INTERAKTIONEN GIBT ES AUCH ASYMMETRISCHE VERTEILUNG VON FÄHIGKEITEN.

All dies wäre nicht möglich gewesen, wenn ich nicht die Möglichkeit gehabt hätte, mit Menschen und Maschinen zu interagieren.



Daniel
Dennett



Eric
Schwitzgebel



Mathew
Crosby



David
Schwitzgebel



Mike
Wilby



DigiDan

Vielen Dank für
Ihre
Aufmerksamkeit!



A HUMAN-MADE BOOK IN THE AGE OF MACHINE-GENERATED TEXTS

Anna's AI Anthology
How to live with smart machines?

With the release of ChatGPT, large language models (LLMs) have brought us a new form of generational shift and scientific debate.

The publication of this book is a first in itself!

Can philosophical considerations help us to work out how we can live with such smart machines? In this book, distinguished philosophers explore questions such as whether these new machines are able to act, whether they are moral agents, whether they have communicative skills, and if they might have relevant concerns.

The book includes contributions from:

- David Gunkel
- David Gunkel
- Thomas Inge
- Julian Nida-Rümelin
- Anna Schwitzgebel
- Michael Veale

- Christoph Neuhoff
- Thomas Inge
- Eric Schwitzgebel
- Anna Schwitzgebel
- Michael Veale

- Stephan Schulz
- Stephan Schulz
- Henry Grell
- Michael Veale

Edited by Anna & Emily Thomson

xenomei

References

- Agrawal, A., Mackey, L., & Kalai, A. T. (2023). *Do Language Models Know When They're Hallucinating References?* (arXiv:2305.18248). arXiv. <http://arxiv.org/abs/2305.18248>
- Agüera y Arcas, B. (2022). Do Large Language Models Understand Us? *Daedalus*, 151(2), 183–197. https://doi.org/10.1162/daed_a_01909
- Barkham, P. (2021, July 25). Should rivers have the same rights as people? *The Guardian*. <https://www.theguardian.com/environment/2021/jul/25/rivers-around-the-world-rivers-are-gaining-the-same-legal-rights-as-people>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Bunten, A., Iorns, C., Townsend, J., & Borrows, L. (2021, June 3). *Rights for nature: How granting a river 'personhood' could help protect it*. The Conversation. <http://theconversation.com/rights-for-nature-how-granting-a-river-personhood-could-help-protect-it-157117>
- Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Cole, S. (2023). 'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection. *Vice*. <https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates>
- Dooley, B., & Ueno, H. (2022, April 24). This Man Married a Fictional Character. He'd Like You to Hear Him Out. *The New York Times*. <https://www.nytimes.com/2022/04/24/business/akihiko-kondo-fictional-character-relationships.html>
- Gunkel, D. J. (2023). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. The MIT Press. <https://doi.org/10.7551/mitpress/14983.001.0001>
- Henrich, J. P. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, 17(5), 647–659. <https://doi.org/10.1111/desc.12148>
- (2015). Animal mindreading: What's the problem? *Psychonomic Bulletin & Review*, 22(2), 313–327. <https://doi.org/10.3758/s13423-014-0704-4>
- Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies. *Theory, Culture & Society*, 30(2), 84–115. <https://doi.org/10.1177/0263276412456564>
- Lake, B. M., & Baroni, M. (2023). Human-like Systematic Generalization through a Meta-learning Neural Network. *Nature*, 1–7. <https://doi.org/10.1038/s41586-023-06668-3>
- Lemoine, B. (2022, June 11). Is LaMDA Sentient? — An Interview. *Medium*. <https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d916d917>
- Marcus, G., & Davis, E. (2020). GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about. *MIT Technology Review*. <https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion>
- Michael, J., Sebanz, N., & Knoblich, G. (2016). The Sense of Commitment: A Minimal Approach. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01968>
- Michael, J., Holtzman, A., Parrish, A., Mueller, A., Wang, A., Chen, A., Madaan, D., Nangia, N., Pang, R. Y., Phang, J., & Bowman, S. R. (2022). What Do NLP Researchers Believe? Results of the NLP Community Metasurvey (arXiv:2208.12852).
-

References

- O'Donnell, E., & Talbot-Jones, J. (2017, March 23). Three rivers are now legally people – but that's just the start of looking after them. *The Conversation*.
<http://theconversation.com/three-rivers-are-now-legally-people-but-thats-just-the-start-of-looking-after-them-74983>
- Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese*, 190(10), 1817–1839. <https://doi.org/10.1007/s11229-013-0263-7>
- Perner, J. (1991). *Understanding the representational mind* (pp. xiv, 348). The MIT Press.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526. <https://doi.org/10.1017/S0140525X00076512>
- Robertson, J. (2014). Human Rights vs. Robot Rights: Forecasts from Japan. *Critical Asian Studies*, 46(4), 571–598. <https://doi.org/10.1080/14672715.2014.960707>
- (2017). *Robo sapiens japonicus: Robots, Gender, Family, and the Japanese Nation*.
- Salmón, E. (2000). Kincentric Ecology: Indigenous Perceptions of the Human-Nature Relationship. *Ecological Applications*, 10(5), 1327–1332. <https://doi.org/10.2307/2641288>
- Scarborough, J. K., & Bailenson, J. N. (2014). Avatar Psychology. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality*. Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199826162.013.033>
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. The MIT Press. <https://doi.org/10.7551/mitpress/9780262016797.001.0001>
- Strasser, A. (2006). *Kognition künstlicher Systeme*: DE GRUYTER. <https://doi.org/10.1515/9783110321104>
- (2013). *Kognition künstlicher Systeme*. In *Kognition künstlicher Systeme*. De Gruyter. <https://doi.org/10.1515/9783110321104>
- (Ed.). (2024). *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.
- Strasser, A., & Schwitzgebel, E. (2024). Quasi-sociality: Toward Asymmetric Joint Actions. In *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.
- Strasser, A., & Wilby, M. (2023). The AI-Stance: Crossing the Terra Incognita of Human-Machine Interactions? In *Social Robots in Social Institutions* (pp. 286–295). IOS Press.
<https://doi.org/10.3233/FAIA220628>
- Tomasello, M. (2008). *Origins of human communication* (pp. xiii, 393). MIT Press.
- Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models Know What Humans Know? *Cognitive Science*, 47(7), e13309.
<https://doi.org/10.1111/cogs.13309>
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8), 998–1003.
<https://doi.org/10.1016/j.neunet.2010.06.002>
- Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative Activities in Young Children and Chimpanzees. *Child Development*, 77(3), 640–663. <https://doi.org/10.1111/j.1467-8624.2006.00895.x>
- Weil, E. (2023, March 1). *You Are Not a Parrot*. New York Magazine. <https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html>
- Wilby, M., & Strasser, A. (2024). Situating machines within normative practices: Bridging responsibility gaps with the AI-Stance. In A. Strasser (Ed.), *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.