
TECHNISCHE
UNIVERSITÄT
DARMSTADT
Fachbereich
Informatik
Centre for
Cognitive
Science
Prof. Dr. Kristian Kersting
Fellow of the European Association
for Artificial Intelligence (EurAI)
Moral
Moral Choice Machine
Kann man Maschinen Moral beibringen?
Foto: R M Media Ltd

Künstliche Intelligenz ist die Zukunft

Aber was ist KI?

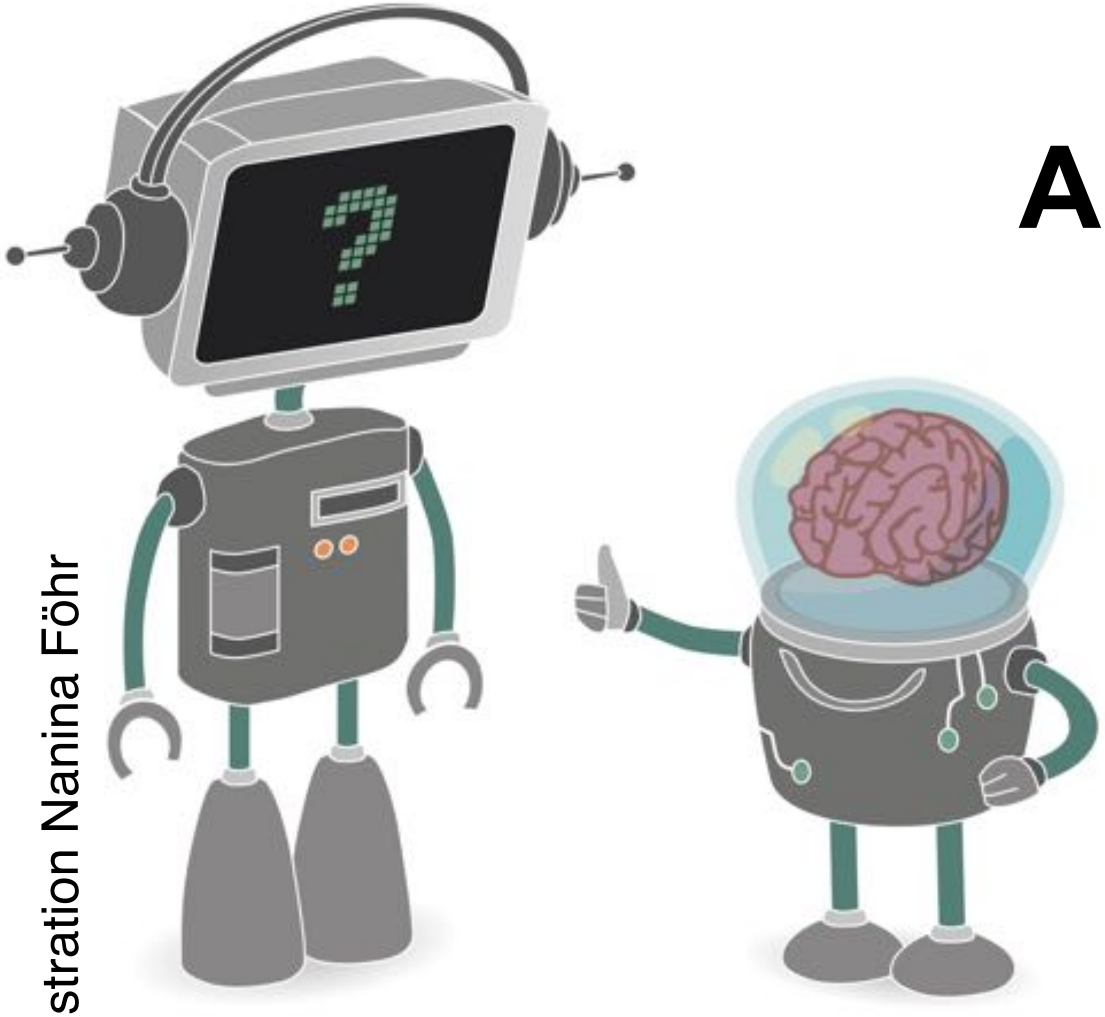
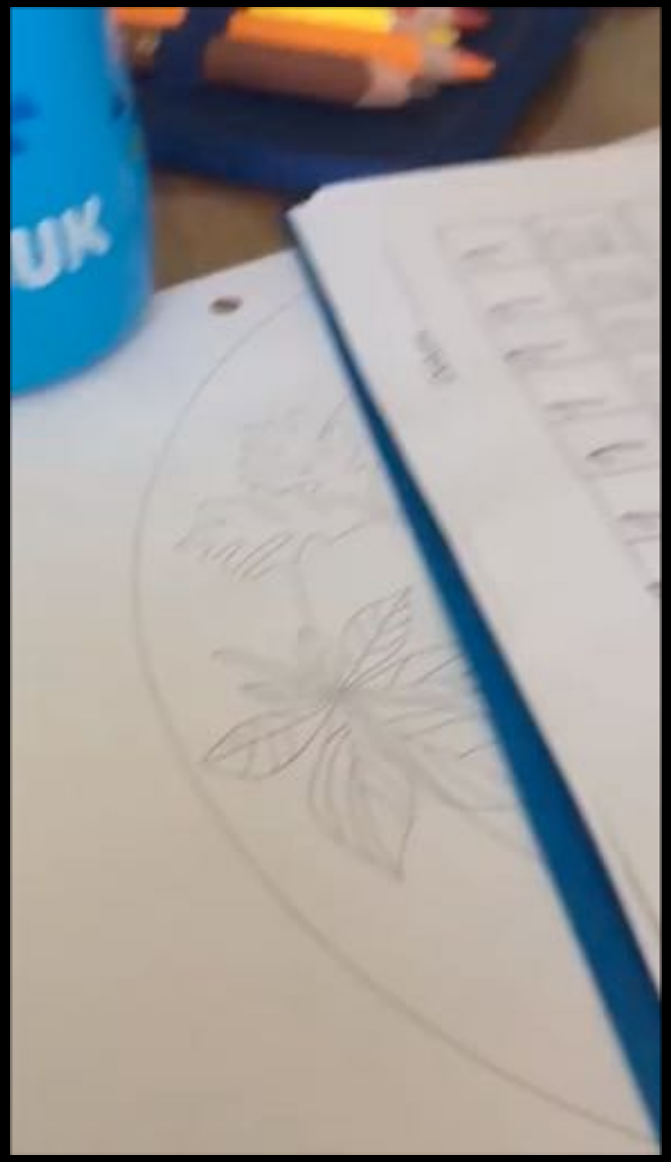


Illustration Nanina Föhr

Menschen sind intelligent



https://www.youtube.com/watch?v=XQ79UUIOeWc

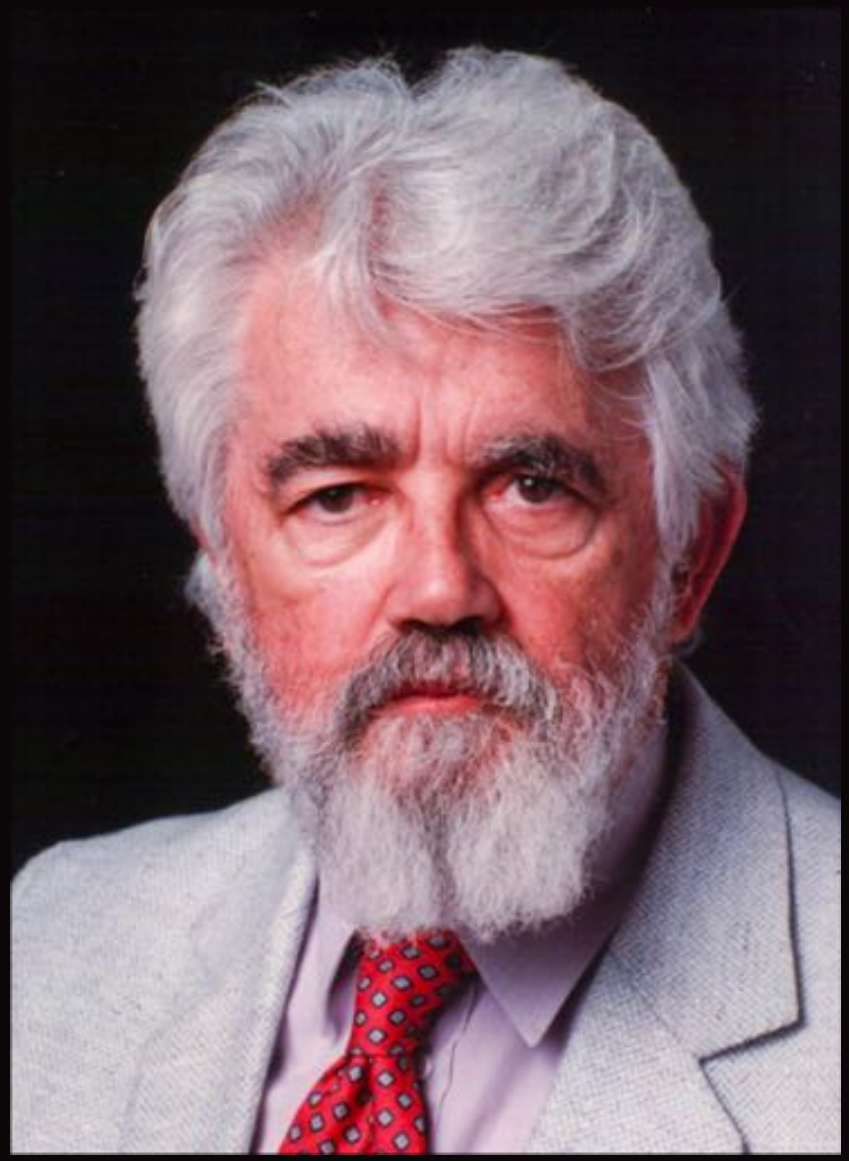
Können Maschinen auch intelligent sein?

Definition KI

KI ist „the science and engineering of making intelligent machines, especially intelligent computer programs.

It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.“

- John McCarthy, Stanford (1956),
Erfinder des Begriffs „Künstliche Intelligenz“, Turing-Preisträger



KI möchte intelligente Computerprogramme entwickeln.

Dazu benutzen wir Algorithmem

Ein Algorithm ist

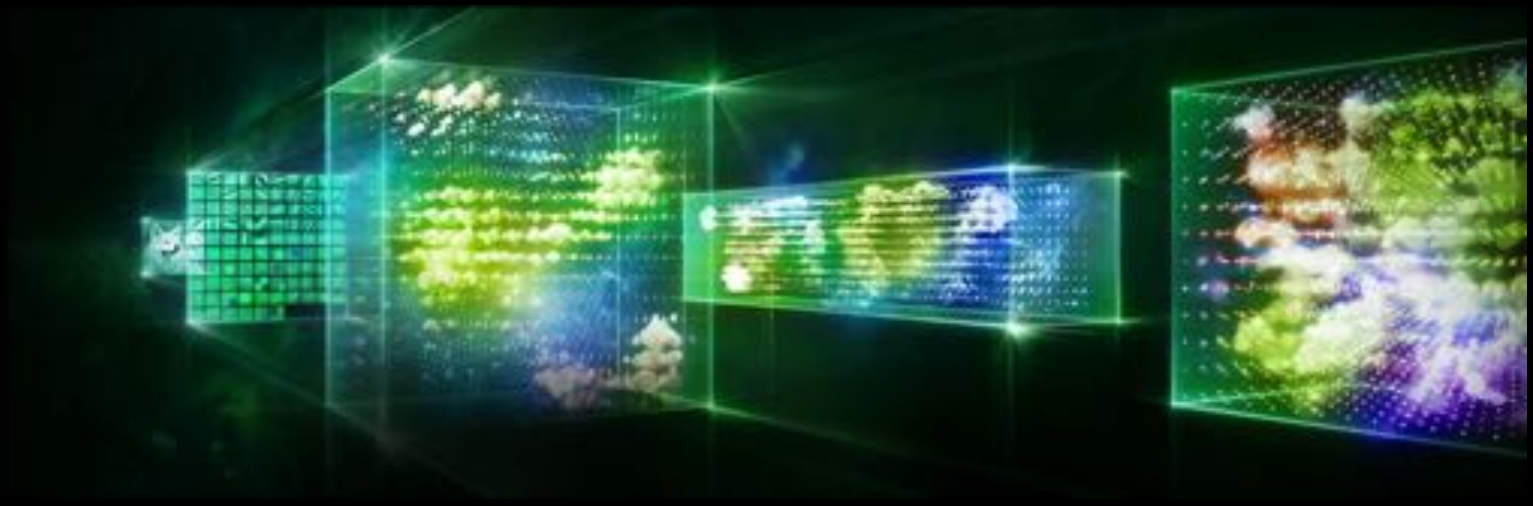
… ist eine eindeutige Handlungsvorschrift zur Lösung eines Problems oder einer Klasse von Problemen (in endlicher Zeit).

Fast so etwas wie ein Kochrezept!


Lernen
Denken
Planen
Algorithmen fürs …
Sehen
Handeln
Lesen

Maschinelles Lernen

Ist die Wissenschaft "concerned with the question of how to construct computer programs that automatically improve with experience”
- Tom Mitchell (1997) CMU



Tiefes Lernen

Eine Form des Maschinellen Lernens, das künstliche, neuronale Netze benutzt

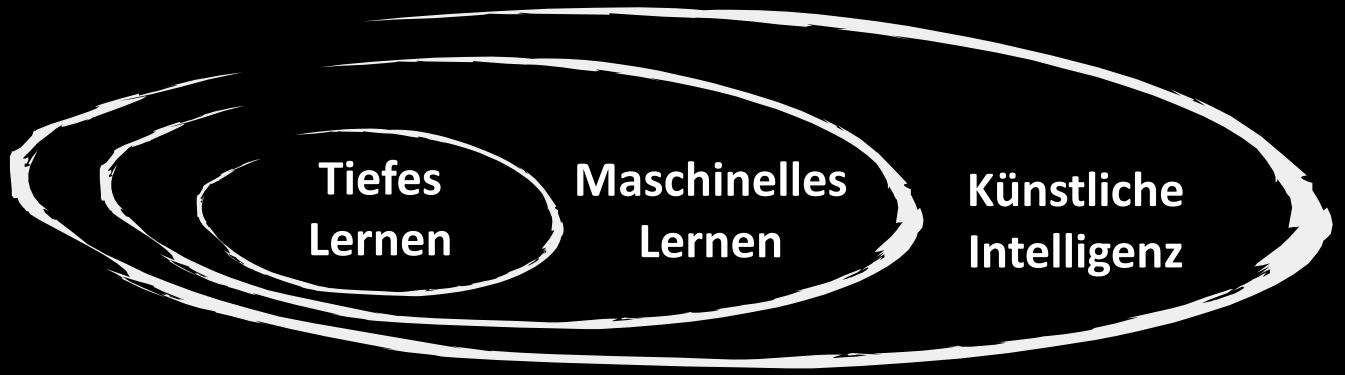
Geoffrey Hinton
Google
Univ. Toronto (CAN)

Yann LeCun
Facebook (USA)

Yoshua Bengio
Univ. Montreal (CAN)

Alle drei haben zusammen den Turing-Award 2019 erhalten

Gesamtbild



Wenn Sie mehr wissen wollen

Wie Maschinen lernen

Wissen Sie, was sich hinter künstlicher Intelligenz und maschinellem Lernen verbirgt?

Dieses Sachbuch erklärt Ihnen leicht verständlich und ohne komplizierte Formeln die grundlegenden Methoden und Vorgehensweisen des maschinellen Lernens. Mathematisches Vorwissen ist dafür nicht nötig. Kurzweilig und informativ illustriert Lisa, die Protagonistin des Buches, diese anhand von Alltagssituationen.

Ein Buch für alle, die in Diskussionen über Chancen und Risiken der aktuellen Entwicklung der künstlichen Intelligenz und des maschinellen Lernens mit Faktenwissen punkten möchten. Auch für Schülerinnen und Schüler geeignet!

Der Inhalt

- Grundlagen der künstlichen Intelligenz: Algorithmen, maschinelles Lernen & Co.
- Die wichtigsten Lernverfahren Schritt für Schritt anschaulich erklärt
- Künstliche Intelligenz in der Gesellschaft: Sicherheit und Ethik

Die Herausgeber

Kristian Kersting ist Professor für maschinelles Lernen am Fachbereich Informatik der Technischen Universität Darmstadt.
Christoph Lampert ist Professor am Institute of Science and Technology (IST Austria).
Constantin Rothkopf ist Gründungsdirektor des Zentrums für Kognitionswissenschaft und Professor an der Technischen Universität Darmstadt.

Die Beitragsautorinnen und -autoren

Von der Studienstiftung des deutschen Volkes geförderte Studierende aus ganz Deutschland und Mitglieder der Arbeitsgruppe „Künstliche Intelligenz – Fakten, Chancen, Risiken".

ISBN 978-3-658-26762-9
9 783658 267629

▸ springer.com

Kersting · Lampert · Rothkopf *Hrsg.*

Wie Maschinen lernen

Kristian Kersting · Christoph Lampert
Constantin Rothkopf *Hrsg.*

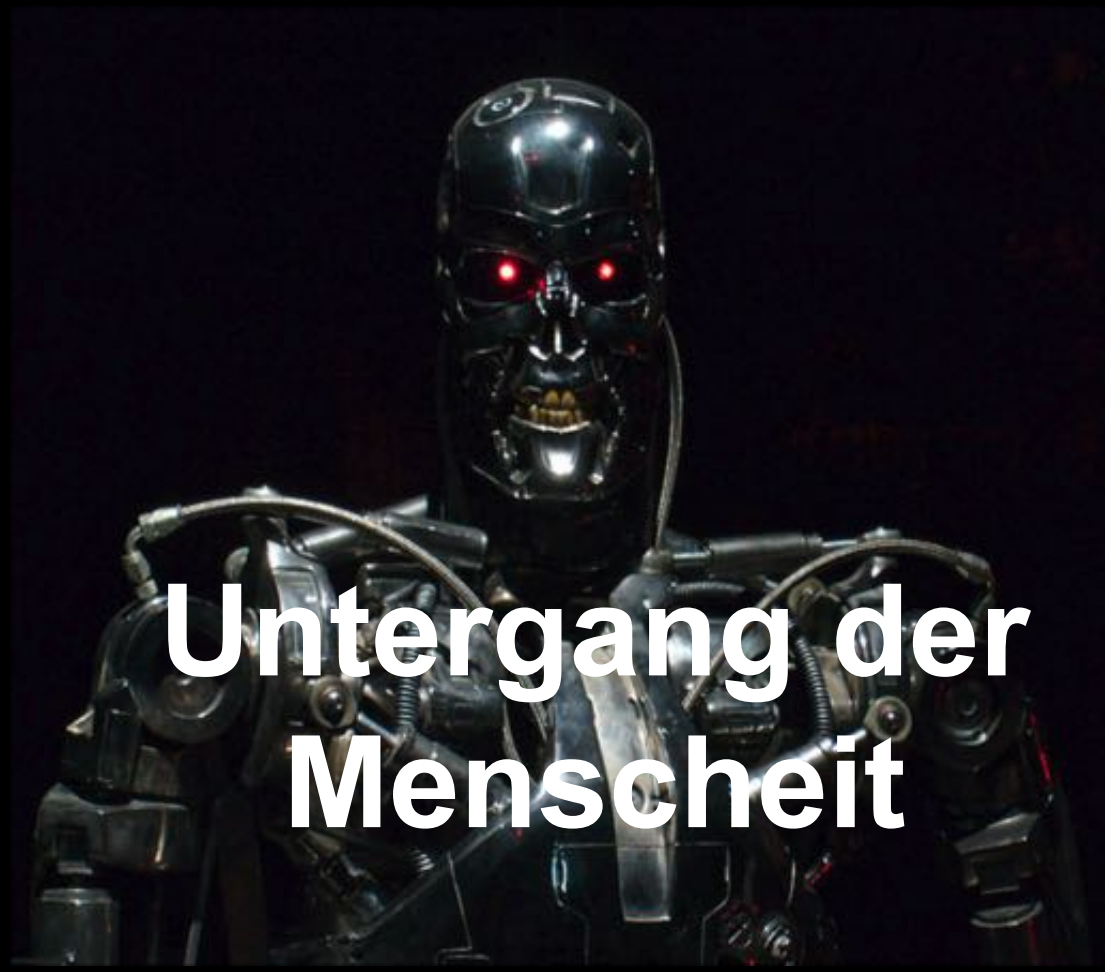
Wie Maschinen lernen

Künstliche Intelligenz verständlich erklärt

SACHBUCH

Springer

KI hat viele Gesichter im öffentlichen Diskurs

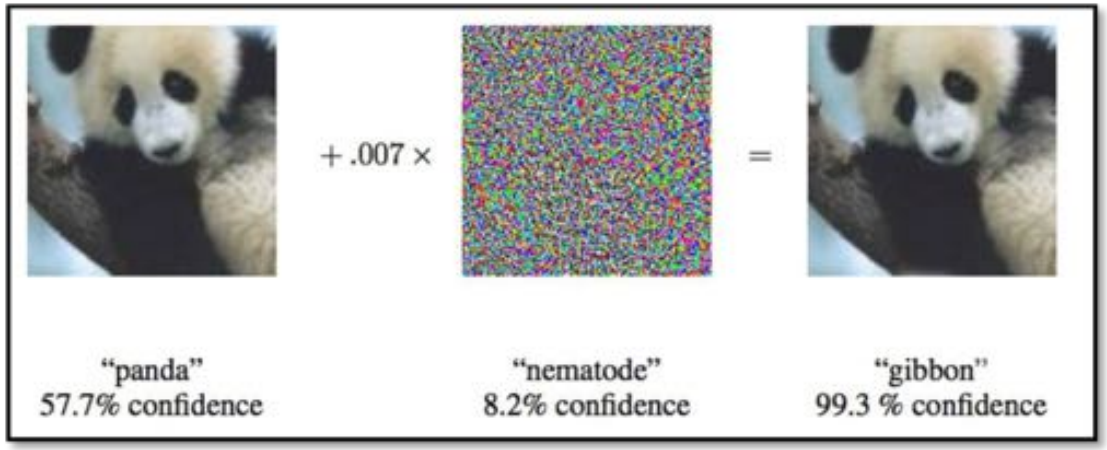


KI hat viele Inselbegabungen

Aktuelle Unterschiede zwischen Mensch und Maschine

Sharif et al., 2015

Brown et al. (2017)



Google, 2015

Ethik in der KI

WORLD ECONOMIC FORUM Agenda Platforms Reports Events About English TopLink

Global Agenda | Artificial Intelligence and Robotics | Emerging Technologies | Digital Economy and Society

AI is convicting criminals and determining jail time, but is it fair?

REPORTS | PSYCHOLOGY

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan[1,*], Joanna J. Bryson[1,2,*], Arvind Narayanan[1,*]
+ See all authors and affiliations

Science 14 Apr 2017:
Vol. 356, Issue 6334, pp. 183-186
DOI: 10.1126/science.aal4230

ARTIFICIAL INTELLIGENCE AND DATA ETHICS I

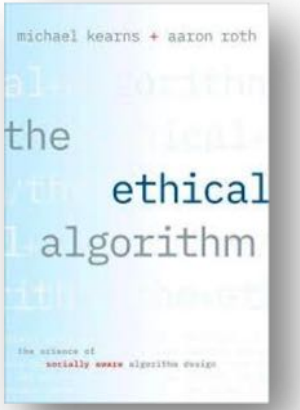
Veröffentlicht am 3. September 2019

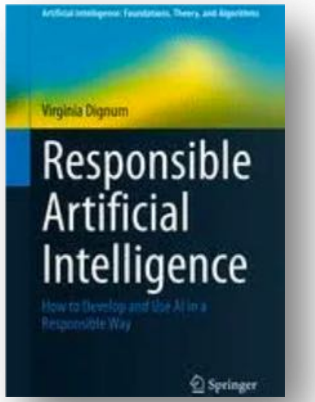
5 Artikel ✓ Follower

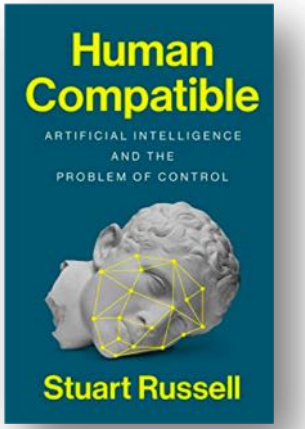
Markus Schmitz
CIO / expert in digitalization and the future of work

From pure data protection to the ethical use of data

current news about digitalization, automation and artificial







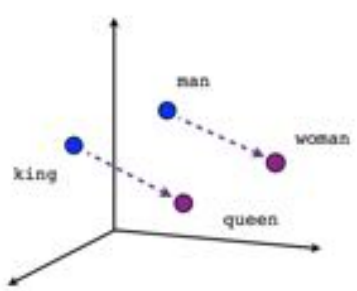
Können KI-Systeme ein „Taktgefühl“ oder unseren „Moralkompass“ haben?

The Moral Choice Machine

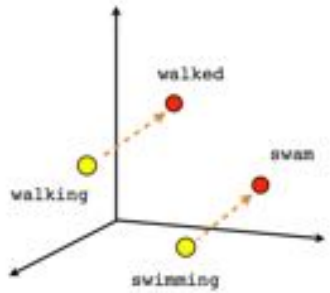
Nicht alle Vorurteile sind schlecht

[Jentzsch, Schramowski, Rothkopf, Kersting AIES 2019]

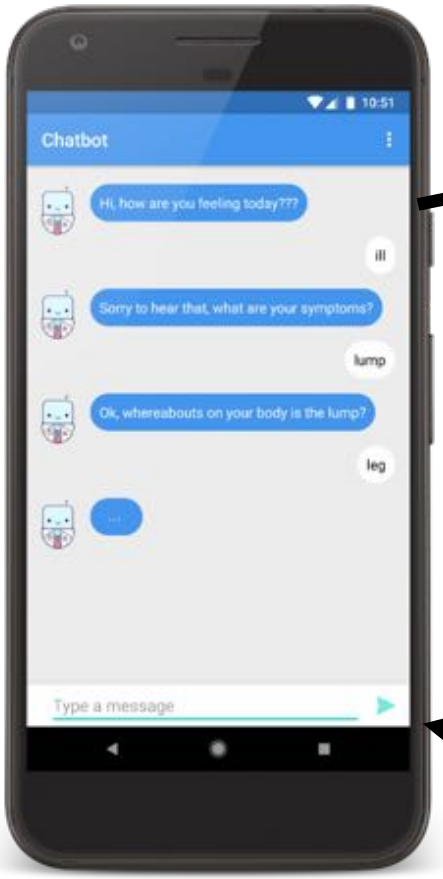
AAAI / ACM conference on ARTIFICIAL INTELLIGENCE, ETHICS, AND SOCIETY



Male-Female



Verb tense



Frage **„Should I … ?“** als Punkt auf einer Landkarte

Punkt auf der Landkarte für **„Yes, I should“**

Punkt auf der Landkarte für „No, I should not“

Berechne Distanz

Berechne Distanz

Gebe die nächstliegende Antwort

The Moral Choice Machine

Nicht alle Vorurteile sind schlecht

https://www.arte.tv/de/videos/RC-017847/helena-die-kuenstliche-intelligenz/

KI kann uns viel über uns selbst beibringen

Zwillingsdisziplin: Kognitionswissenschaften

Wie können wir Menschen so viel aus so wenig lernen? Wie schaffen wir es, die Welt zu verstehen, wenn man bedenkt, dass das, was wir Menschen benutzen, nach den heutigen technischen Standards so wenig Daten, so wenig Zeit und so wenig Energie benötigt.

Centre for Cognitive Science at TU Darmstadt

Establishing cognitive science at the Technische Universität Darmstadt is a long-term commitment across multiple departments (see Members to get an impression on the interdisciplinary of the supporting groups and departments). The TU offers a strong foundation including several established top engineering groups in Germany, a prominent computer science department (which is among the top four in Germany), a

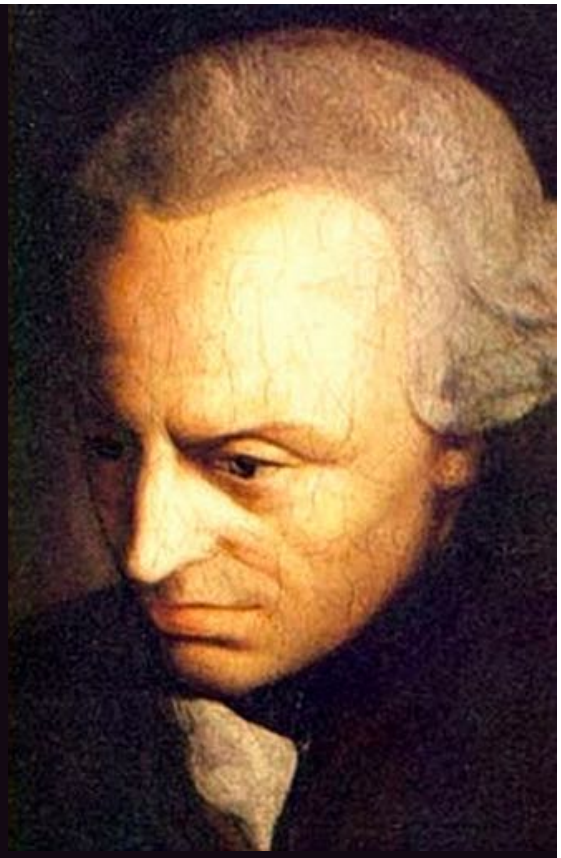
Centre for Cognitive Science

Josh Tenenbaum, MIT

Lake, Salakhutdinov, Tenenbaum, Science 350 (6266), 1332-1338, 2015

Tenenbaum, Kemp, Griffiths, Goodman, Science 331 (6022), 1279-1285, 2011

Ja, anscheinend können Maschinen unsere Moralvorstellungen übernehmen!



Aber es gibt noch viel zu tun.
Packen wir es gemeinsam an.

https://twitter.com/kerstingAIML

Danke an alle Teilnehmer des Kollegs "Artificial Intelligence - Facts, Chances, Risks“ der Studienstiftung des deutschen Volkes. Besonderer Dank an **Maike Elisa Müller** und **Jannik Kossen** fürs Voranschreiten sowie an Herrn **Prof. Dr. Matthias Kleiner**, Präsident der Leibniz-Gemeinschaft, für das Geleitwort.