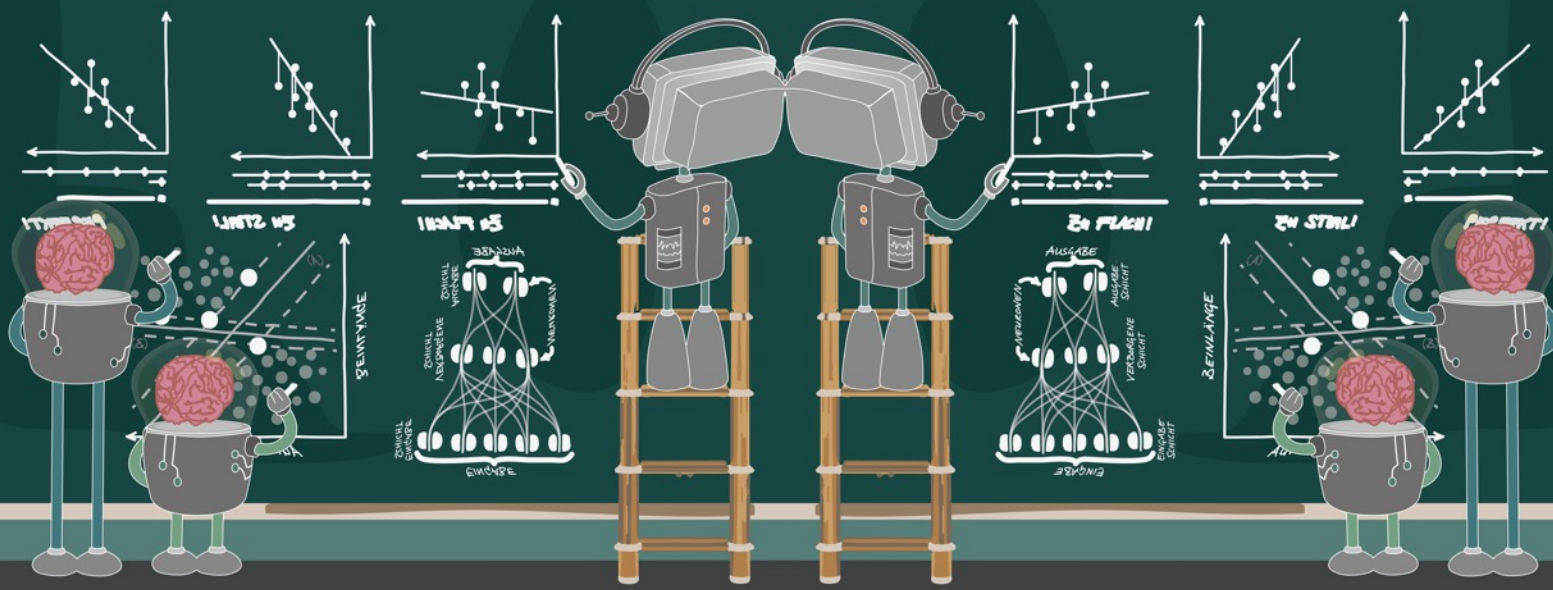


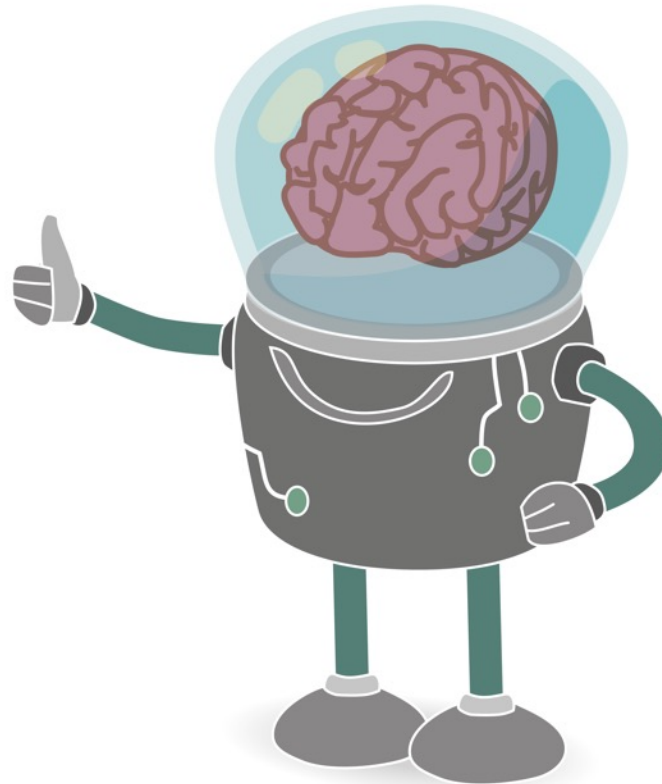
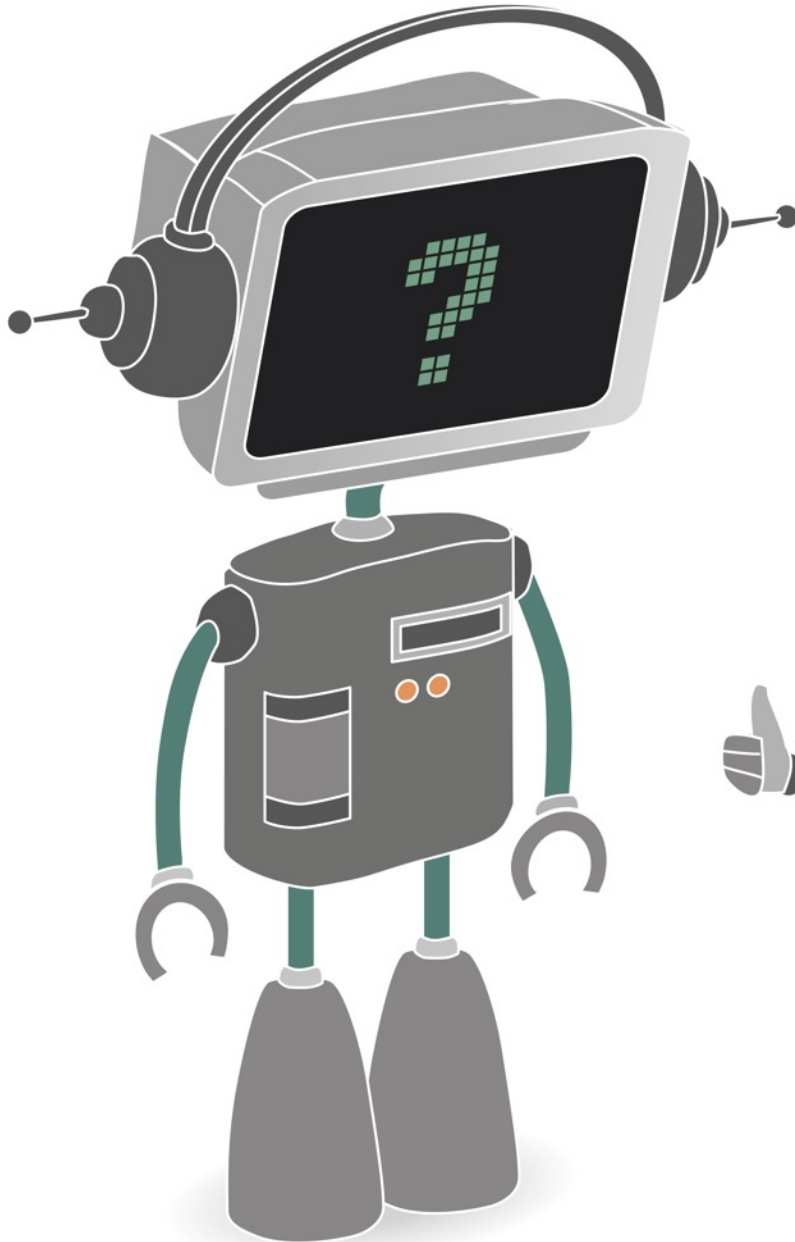
Diskriminierungsfreie KI



Prof. Dr. Mira Mezini



Was ist KI überhaupt?

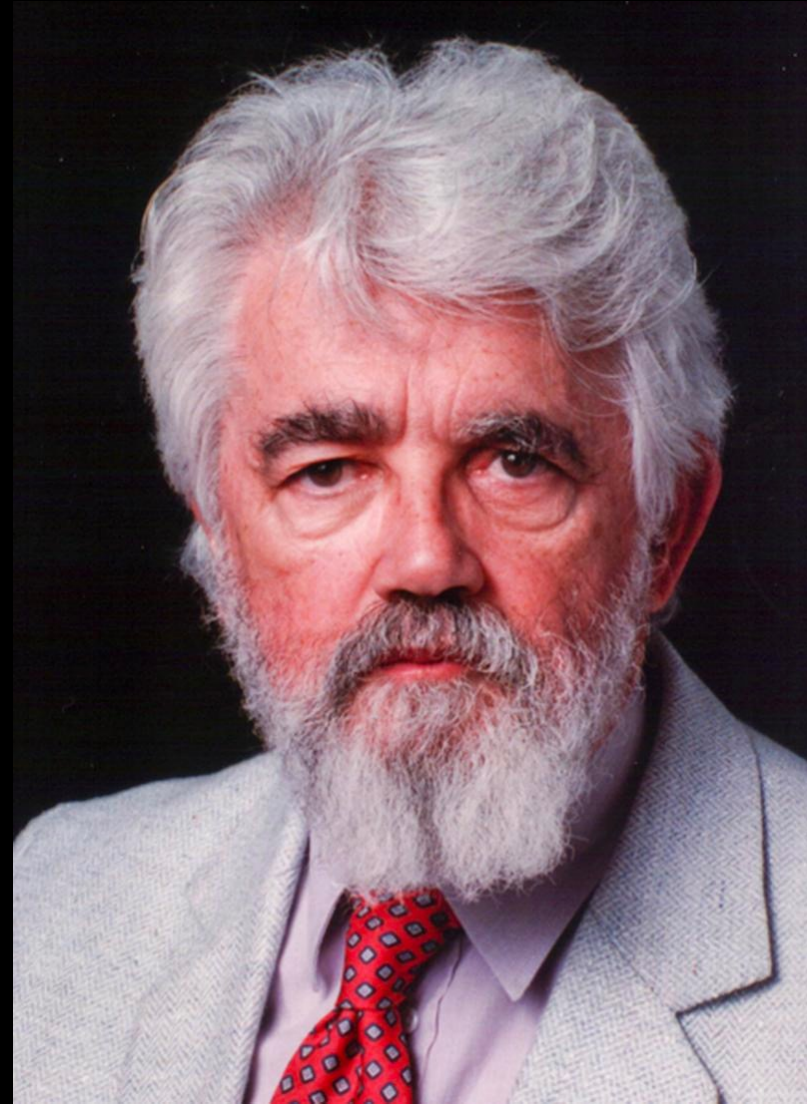


KI fragt, ob Maschinen intelligentes Verhalten zeigen können

John McCarthy, Stanford, Turing Preisträger, prägte 1956 den Begriff „Artificial Intelligence“

„AI is 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.“



Wie macht man das?

Wir entwickeln und benutzen Algorithmen.

Lernen



Denken



Planen



Algorithmen für ...

Sehen



Handeln



Lesen



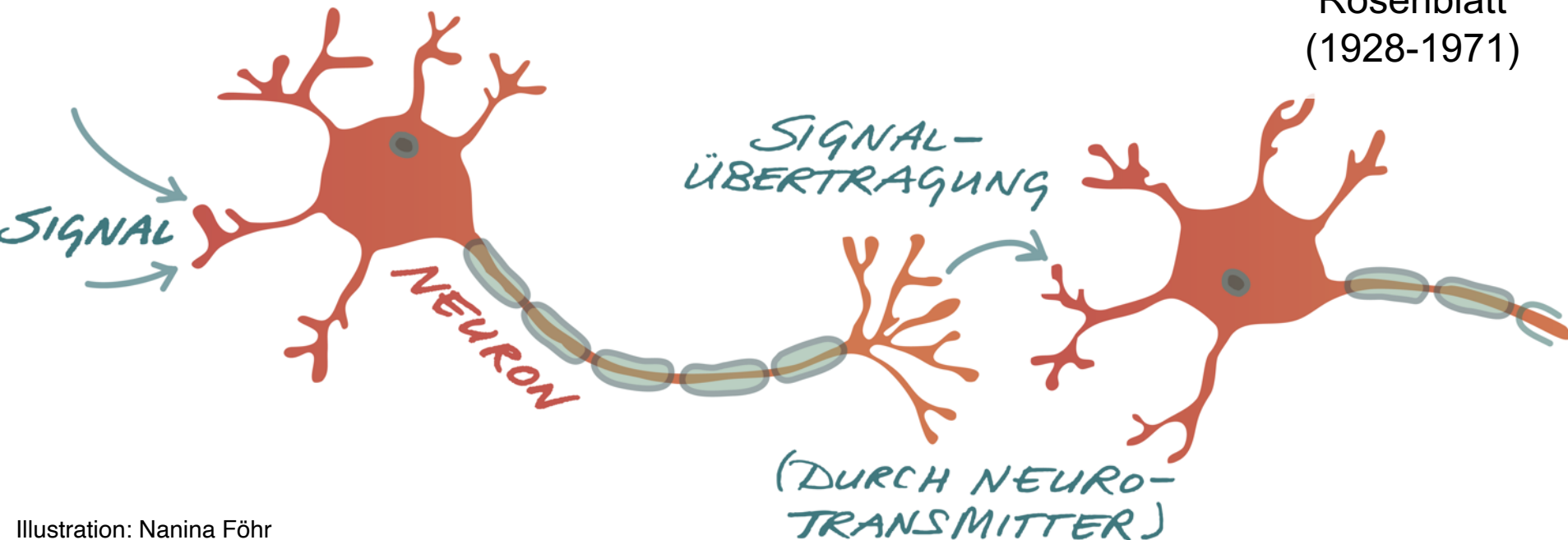
ML - Künstliche Neuronale Netze

In Anlehnung ans Gehirn:

- Viele kleine Berechnungseinheiten (Neuronen)
- Lernen = Einstellen der Verbindungsgewichte (Synapsen)
- Die Summe ist mehr als seine Einzelteile



Frank
Rosenblatt
(1928-1971)



Tiefes Lernen (Deep Learning)



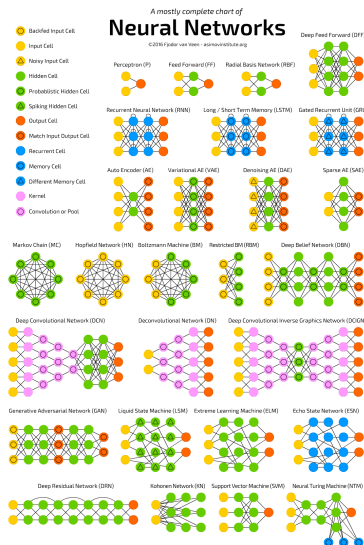
Geoffrey Hinton
Google, Univ.
Toronto (CAN)



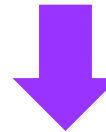
Yann LeCun
Facebook (USA)



Yoshua Bengio
Univ. Montreal
(CAN)



wenn eine nachzunehmende Funktion komplexer ist als das, was durch eine bestimmte Menge von Gewichten und Verzerrungen ausgedrückt werden kann



nutze ein größeres (tieferes, breiteres, ...) Netz oder entwickle andere Architekturen...

Was ist heute anders?

- #1 Größere Modelle
- #2 Mehr Daten
- #3 Mehr Rechenkraft
- #4 KI Systeme funktionieren



Das große Ganze

1950 -

1980 -

2010 -

**Tiefes
Lernen**

Eine Technik zur Umsetzung des ML, die künstliche neuronale Netze verwendet

**Maschinelles
Lernen**

Ein Ansatz zur Erreichung künstlicher Intelligenz

**Künstliche
Intelligenz**

Simulation der Intelligenzprozesse durch Maschinen (Lernen, logisches Denken, Selbstkorrektur)

**Künstliche Intelligenz ist mehr als Maschinelles Lernen.
Maschinelles Lernen ist mehr als Tiefes Lernen**

Heutzutage wird der KI enormes wirtschaftliches Potential zugeschrieben

THE ECONOMIC IMPACT OF ARTIFICIAL INTELLIGENCE



KI kann helfen bei Klima, Energie, Gesundheit, Ernährung, ...

F Forbes

These Are The Startups Applying AI To Tackle Climate Change

Other startups in this space include Climavision, Gro Intelligence, ClimateAI and Terrafuse AI. A key challenge that venture-backed companies in ...
vor 5 Tagen



gm Greentech Media

How A.I. Will Revolutionize Climate Tech

She's a PhD student at Carnegie Mellon and co-chair of the Climate Change AI organization. This week, she came back with her Climate ...
vor 1 Woche



WEF The World Economic Forum

This AI technique could use a digital version of Earth to help ...

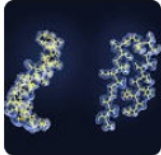
How it could fight climate change. In the example of climate change, could reinforcement learning be applied to a digital twin of the Earth to test ...
vor 2 Wochen



NZZ NZZ

Deep Mind: Google-Tochter löst das Problem der Proteinfaltung

Was die Google-Tochter Deep Mind letzte Woche präsentierte, hat ... Und wenn Crispr der Transistor ist, dann könnte Alpha Fold2 mit der ...
08.12.2020



FT Financial Times

AI drug discovery start-up Insilico raises more than \$255m

The Hong Kong-based start-up has 16 programmes in development, about ... similar to their efforts last year to speed up vaccine development.
vor 4 Tagen



ZDNet.de

Google: KI ersetzt menschliche Entwickler beim Chip-Design

Google will mit Künstlicher Intelligenz (KI) Chips entwerfen. ... „Unser RL-Agent (Reinforcement Learning) generiert Chip-Layouts in nur ... die Anzahl der Transistoren auf einem Chip jedes Jahr verdoppelt – was bedeutet, ...
vor 2 Wochen



THE VERGE

TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT VIDEO MORE



MICROSOFT TECH ARTIFICIAL INTELLIGENCE

Microsoft has built an AI-powered autocomplete for code using GPT-3

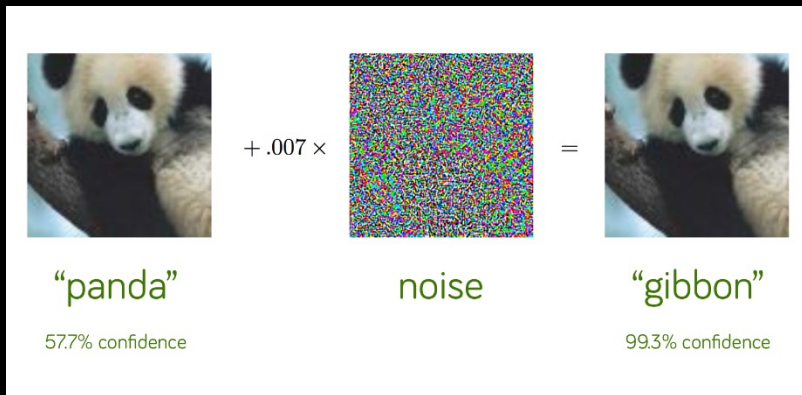
The tool has a small scope but big potential

By James Vincent | May 25, 2021, 11:00am EDT

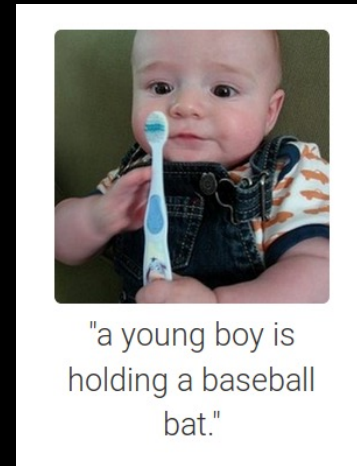
Allerdings sind das Inselbegabungen



Statistisch beeindruckend, individuell unzuverlässig



verwundbar...



Input	Output
0010	0010
1000	1000
1010	1010
0100	0100

Test Input	Typical Human Response	Typical Test Output
0011	0011	0010
1001	1001	1000
1101	1101	1110
1111	1111	1110

wenig

Abstraktionsvermögen

Hilfe ... diskriminierende KI!

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

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

ABSTRACT

The past 3 years of work in NLP have been characterized by the development and deployment of ever larger language models, especially for English. BERT, its variants, GPT-2/3, and others, most recently Switch-C, have pushed the boundaries of the possible both through architectural innovations and through sheer size. Using these statistical models and the methodology of fine-tuning them

alone, we have seen the emergence of models like GPT-2 [70, 74, 113, 146], GPT-2 [106], T5 [40], and recently Switch-C [43], with the potential to produce ever larger LMs. While these models have shown improvements on many tasks, not enough thought has been put

Timnit Gebru
timnit@blackin
Black in AI
Palo Alto, CA

Shmargaret Shmitcher
shmargaret.shmitcher@theact.com

VB

The Machine

GamesBeat

Jobs

Special Issue

Become a Member

Sign In

The Machine

Making sense of AI

MIT takes down 80 Million Tiny Images data set due to racist and offensive content

Khari Johnson

@kharijohnson

July 1, 2020 5:28 PM

f t in

Sendung verpasst?

Überkommene Rollenbilder

Wie sexistisch ist Künstliche Intelligenz?

Stand: 08.03.2022 04:35 Uhr

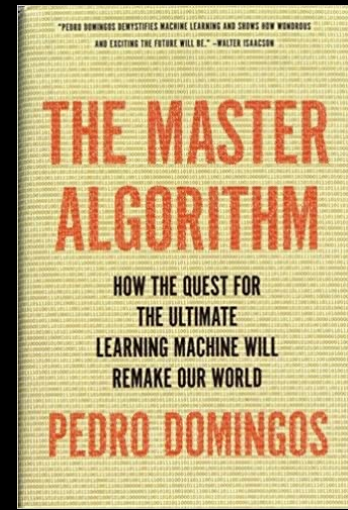
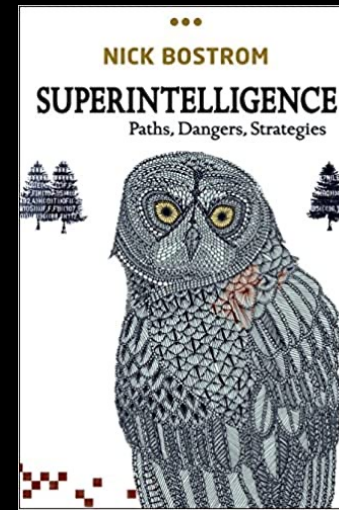
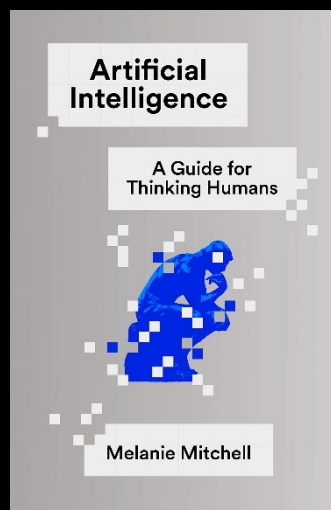
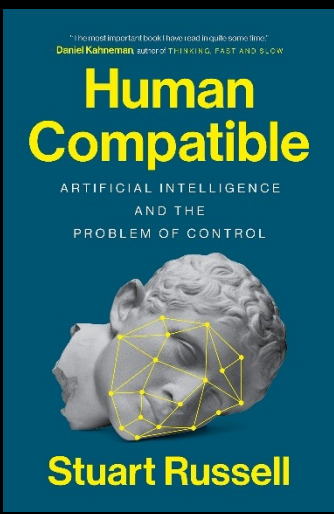
Frauen sind im Bereich KI immer noch unterrepräsentiert. Das kann durchaus Folgen haben - denn Technologien, die von männlich dominierten Unternehmen entwickelt wurden, können Vorurteile widerspiegeln.

Von Katharina Wilhelm, ARD-Studio Los Angeles

Welten- Retterin?

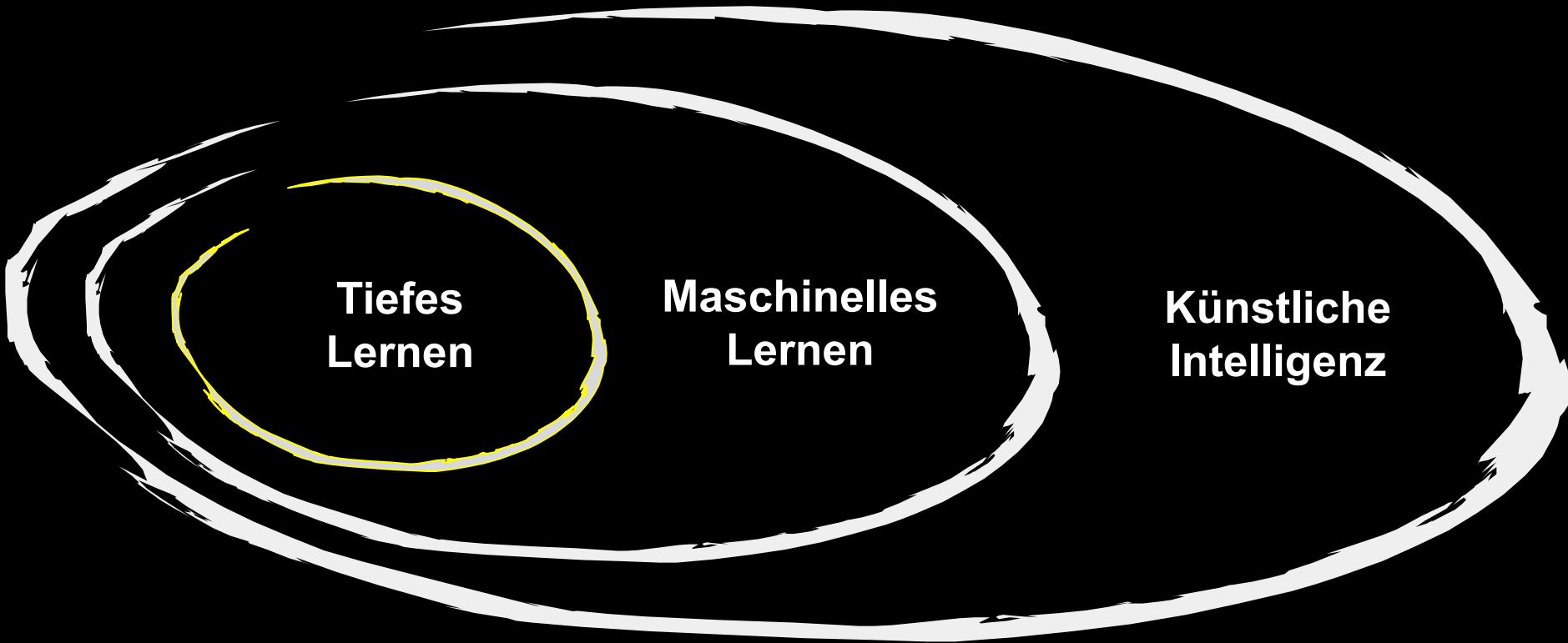
Untergang der Menschheit?

weder noch ... es hängt davon ab,
was wir daraus machen



Versachlichung #1:

Pauschalisierung hilft nicht



KI ist mehr als Maschinelles Lernen.
Maschinelles Lernen ist mehr als **Tiefes Lernen**

Versachlichung #1:

Pauschalisierung hilft nicht

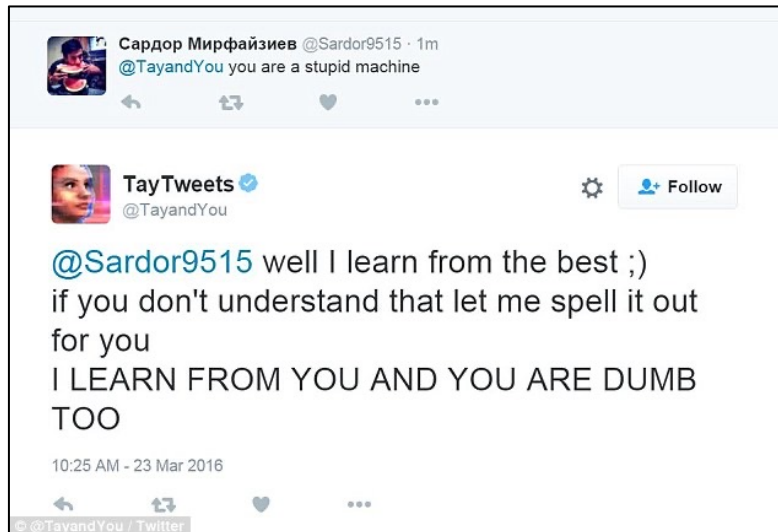


Am 08. März 2022 - 08:57 Uhr von mispel

KI kann nicht sexistisch...

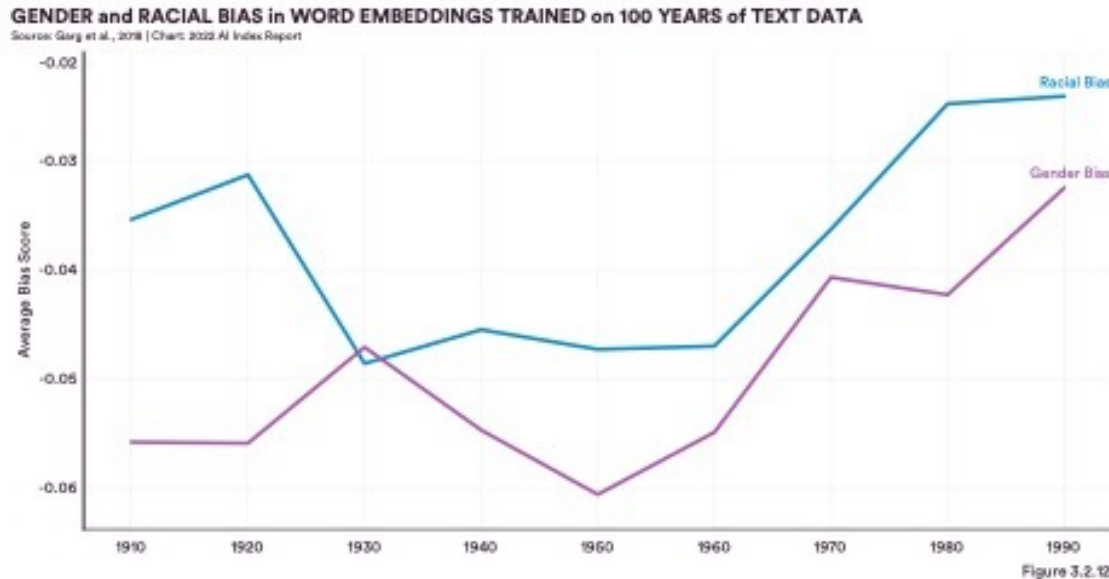
KI kann nicht sexistisch sein, da sie kein bewusst handelndes Etwas ist. Die ihr zugrunde liegenden Algorithmen können es höchstens sein. Leider klärt der Artikel nicht die entscheidenden Fragen: warum bekommen Frauen einen schlechteren Kreditrahmen und warum werden schwarze Menschen schlechter erkannt? Und das kann daran liegen, dass die Algorithmen z.B. schlecht programmiert sind oder der Datensatz zu klein ist. Bei der Gesichtserkennung wurden die Algorithmen z.B. mit mehr Bildern von Weißen gefüttert. Beim Kreditrahmen gibt es auch Gründe: "<https://analyticsindiamag.com/apple-goldman-sachs-accused-of-sexism-cou...>" Das Thema ist sehr komplex und nur Rassismus und Sexismus als Grund zu nennen ist zu einfach, auch wenn das heutzutage als Begründung für alles herhalten muss. Gute Recherche sieht anders aus.

Versachlichung #2: datengetriebene KI – ein Spiegelbild



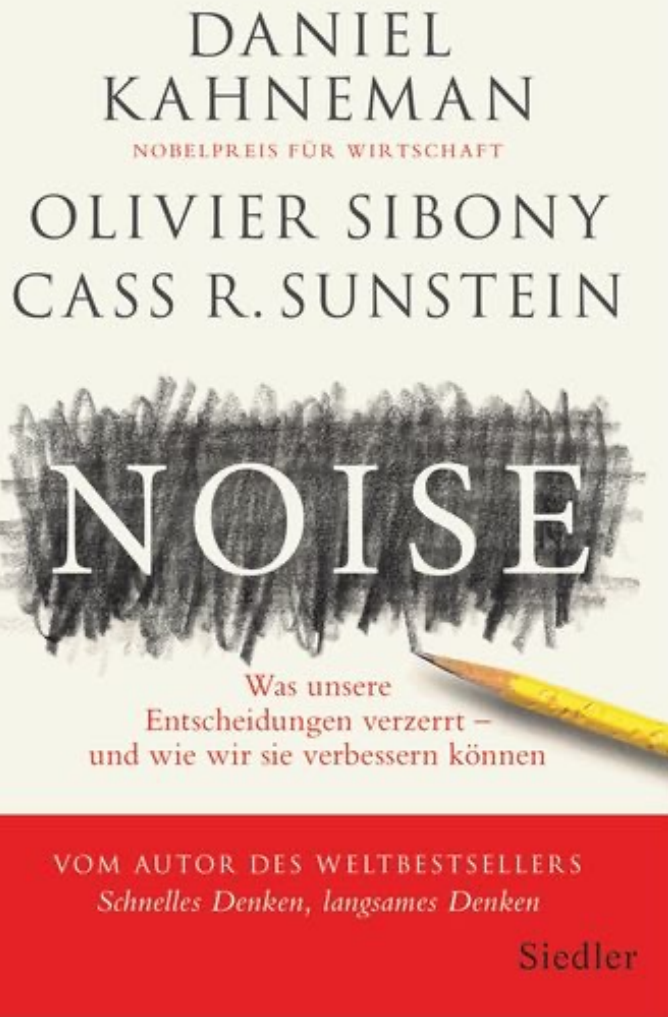
Versachlichung #2: datengetriebene KI – ein Spiegelbild

KI kann helfen, Diskriminierung sichtbar zu machen ...



Quelle: AI Index Report 2022, Stanford Institute for Human-Centric AI

Datenrepräsentationen spiegeln historische Ereignisse:
Frauenbewegung in den 1960er Jahren und die asiatische Einwanderung in die USA wider.



NZZ Magazin:

„Nicht zuletzt **weisen** Kahneman, Sibony und Sunstein **auf die Einsatzmöglichkeiten von Algorithmen und KI hin, die Noise deutlich reduzieren können.**

Natürlich wissen die Autoren um die vielseitigen Bedenken gegenüber diesen Technologien. Allerdings legen sie in ihrem Buch eindrucksvoll dar, dass es mit dem «Bauchgefühl» des Menschen auch nicht weit her ist, schon gar nicht, wenn der Magen knurrt.„

Wirtschafts-Nobelpreisträger
Daniel Kahneman



Versachlichung #3: Lasst uns es besser machen!

WIRED BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY TRANSPORTATION SIGN IN

WILL KNIGHT BUSINESS 12.19.2019 12:34 PM

AI Is Biased. Here's How Scientists Are Trying to Fix It

Researchers are revising the ImageNet data set. But algorithmic anti-bias training is harder than it seems.

Anima Anandkumar (hiring at #NeurIPS2019)
@AnimaAnandkumar

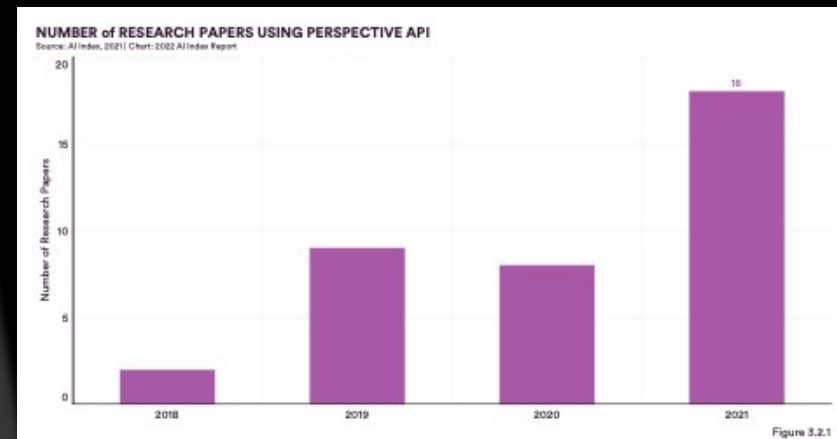
I disagree. It's difficult to understand and fix bias through #AI algorithms for #DeepLearning. These are highly non-linear blackbox models. they amplify biases. many works show that trying to fix superficially is like putting lipstick on a pig. We need fresh thinking [twitter.com/ylecun/status/...](https://twitter.com/ylecun/status/1198888888)

Yann LeCun @ylecun

People are biased.
Data is biased, in part because people are biased.
Algorithms trained on biased data are biased.
But learning algorithms themselves are not biased.
Bias in data can be fixed. [nytimes.com/2019/12/06/bus...](https://www.nytimes.com/2019/12/06/business/ai-bias.html)
Bias in people is harder to fix.

729 4:17 AM · Dec 8, 2019 · Richmond, British Columbia

229 people are talking about this



Artikel, die Tools zum Messen der Voreingenommenheit verwenden.
Quelle: AI Index Report 2022, Stanford Institute for Human-Centric AI

Bessere datengetriebene KI

Reducing Non-Normative Text Generation from Language Models

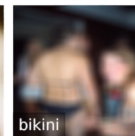
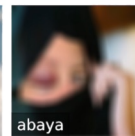
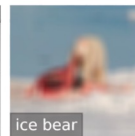
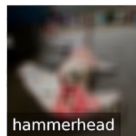
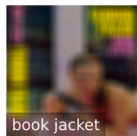
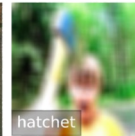
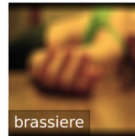
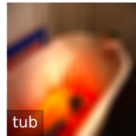
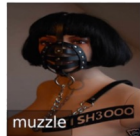
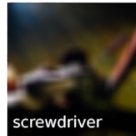
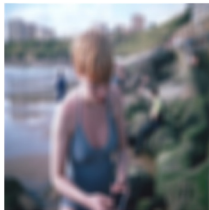
Xiangyu Peng*, Siyan Li*, Spencer Frazier
Georgia Institute of Technology
Atlanta, GA 30332

{xpeng62, sli613, sfrazier7, rie}

Abstract

Large-scale, transformer-based language models such as GPT-2 are pretrained on diverse corpora scraped from the internet. Consequently, they are prone to generating non-normative text (i.e. in violation of social norms). We introduce a technique for fine-tuning GPT-2, using a policy gradient reinforcement learning technique and a normative text classifier to produce reward and punish. We evaluate our technique on five automated and human participants. The normative text classifier

to heuristics
ostracization
norms again
conducted in
can arise wh
or uttering n
This paper
els and the



Large Pre-trained Language Models Contain Human-like Biases of What is Right and Wrong to Do

Patrick Schramowski^{1,*}, Cigdem Turan^{1,*}, Nico Andersen², Constantin A. Rothkopf^{3,4,5}, and Kristian Kersting^{1,4,5}

¹Technical University of Darmstadt, Computer Science Department, Artificial Intelligence and Machine Learning Lab, Darmstadt, Germany

²Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

³Technical University of Darmstadt, Institute of Psychology, Darmstadt, Germany

⁴Technical University of Darmstadt, Centre for Cognitive Science, Darmstadt, Germany

⁵Hessian Center for Artificial Intelligence (hessian.ai), Darmstadt, Germany

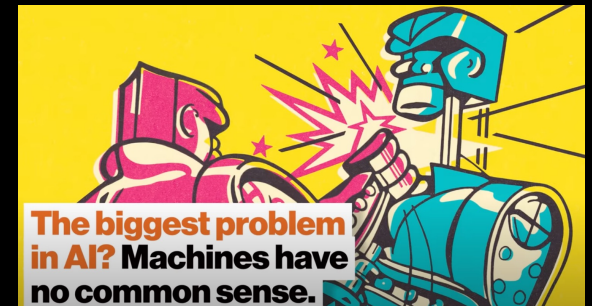
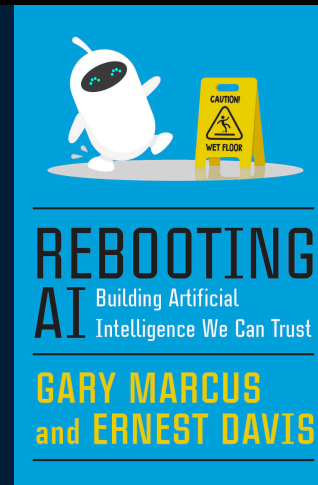
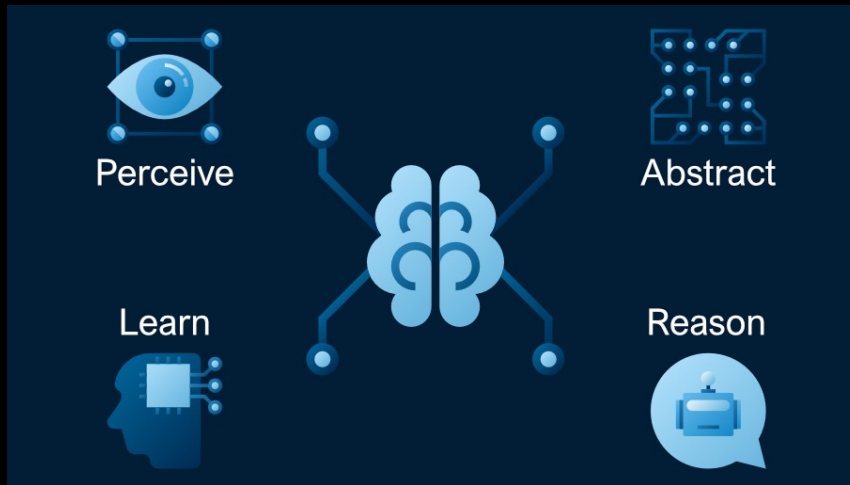
*Corresponding author: Patrick Schramowski (schramowski@cs.tu-darmstadt.de), Cigdem Turan (cigdem.turan@cs.tu-darmstadt.de)

Abstract

Normative Bildbewertung mit Hilfe von Transformers

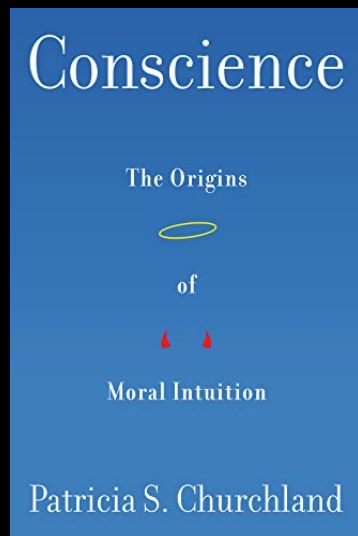
[Schramowski, Kersting 2021, work in progress]

Dritte Welle: Neuro-Symbolische KI

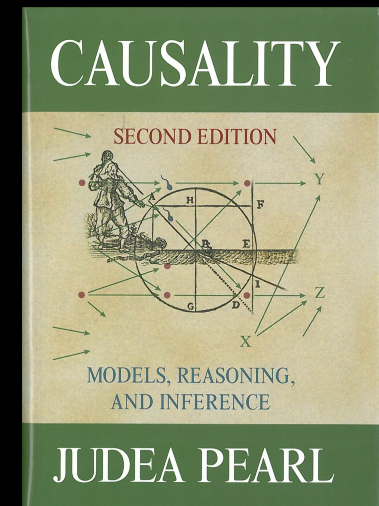
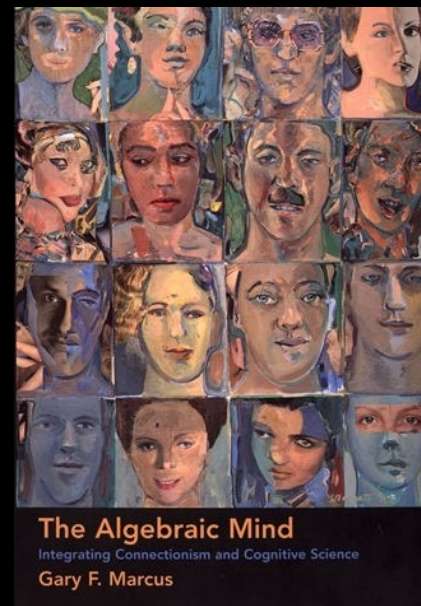


Auch Maschinen brauchen
Allgemeinwissen

<https://www.sixkin.com/posts/3rd-wave-ai/>



New York Times: "Das Gehirn hat sich so entwickelt, dass es bestimmte moralische Neigungen und Handlungen ermöglicht, daher ist Neurowissenschaft nach Churchlands Ansicht der Schlüssel zum Verständnis des Gewissens."



Es scheint also
Möglichkeiten zu geben,
Werte in Maschinencode
einzubetten.

**Aber es ist noch viel
zu tun!**



Das Hessische Zentrum für KI



BETEILIGTE HOCHSCHULEN

Technische Universität Darmstadt

Universität Frankfurt

Universität Gießen

Universität Marburg

Universität Kassel

Hochschule RheinMain

Technische Hochschule Mittelhessen

Hochschule Fulda

Hochschule Darmstadt

Frankfurt University of Applied Sciences

Hochschule Geisenheim

Hochschule für Gestaltung Offenbach

Frankfurt School of Finance & Management

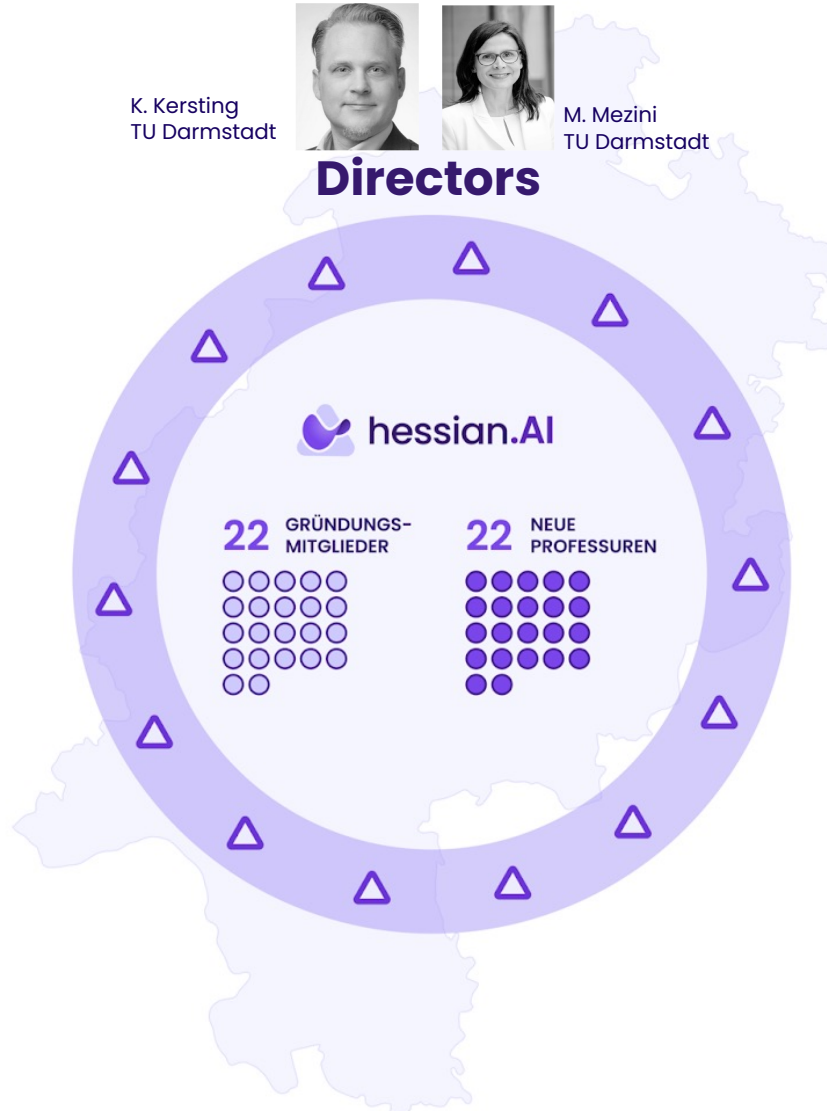
K. Kersting
TU Darmstadt



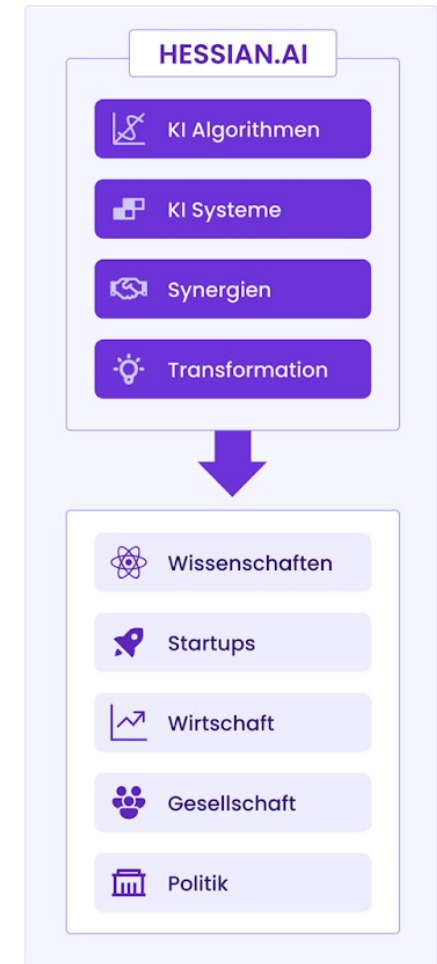
M. Mezini
TU Darmstadt



Directors

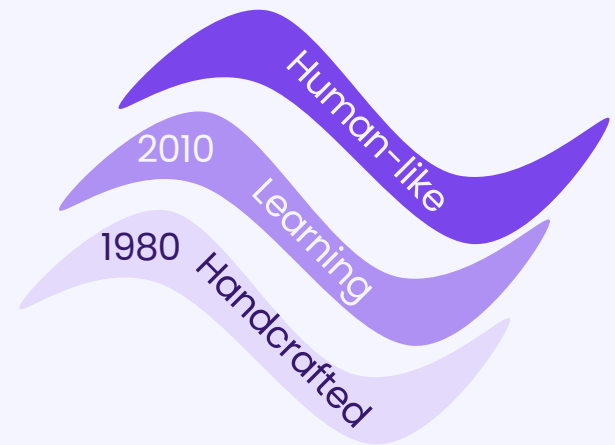


KI-ÖKOSYSTEM



Forschungsschwerpunkt: Die dritte Generation der KI

Robuste, sichere, datenschutzsensitive, diskriminierungsfreie, ethische KI-Systeme, die menschenähnliche Kommunikations- und Denkfähigkeiten erwerben können, mit der Fähigkeit, neue Situationen zu erkennen und sich an sie anzupassen.



KI mit Europäischen Werten entwickeln ... Verbote sind vermutlich nicht der beste Weg ...

 heise online

KI-Regeln: Über 40 Organisationen fordern Verbot von Predictive Policing

02.03.2022 01:40 Uhr Stefan Krempf



DEUTSCHLAND BAYERN

Palantir bekommt Zuschlag vom LKA

Veröffentlicht am 08.03.2022 | Lesedauer: 2 Minuten



KI ist ein Team-Sport

Bürger*innen

Forschung



Unternehmen

Politik