



Ethical
Innovation
Hub



UNIVERSITÄT ZU LÜBECK

Mitbestimmung und KI

Eine Forschungs- und Aktivitätslücke

Prof. Dr.-Ing. Christian Herzog, M.A.

NT & DS Konferenz
18. November 2025





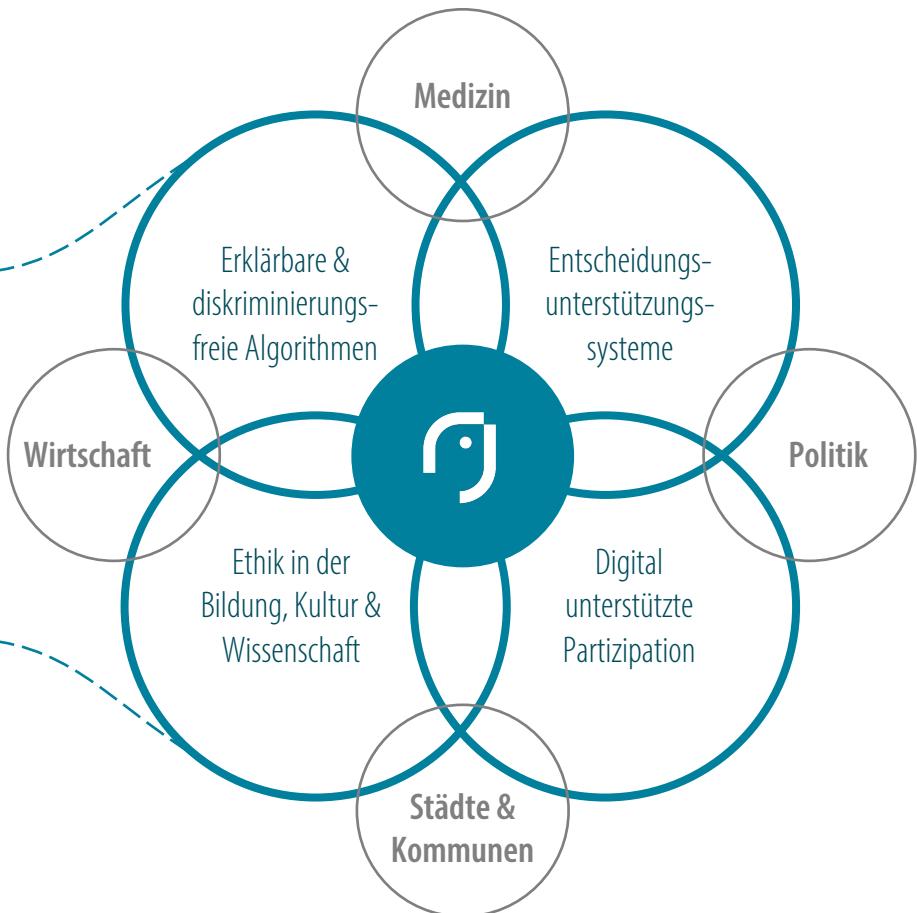
Ethical Innovation Hub

Eine Arbeitsgruppe
zwischen Geistes- &
Technikwissenschaften



INSTITUT FÜR
MEDIZINISCHE
ELEKTROTECHNIK

IMWF



Agenda

KI und Mitbestimmung — Eine Forschungs- und Aktivitätslücke

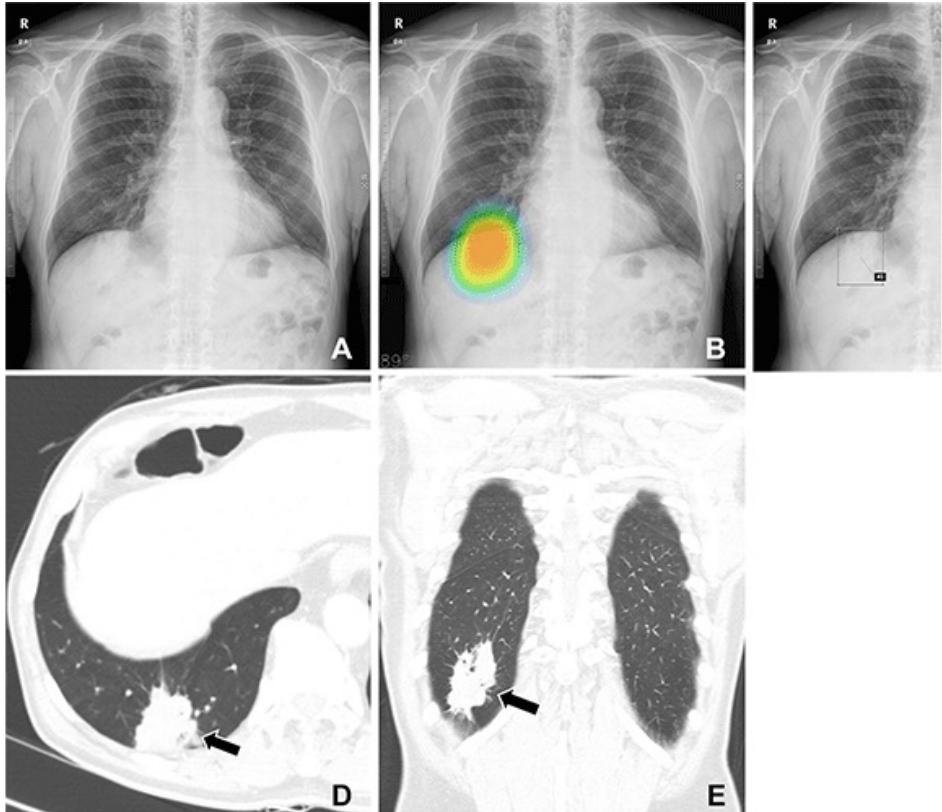
- Was ist Künstliche Intelligenz?
- Aktueller Stand ethischer und rechtlicher Einhegung
- Warum KI eine besondere Herausforderung in der Mitbestimmung ist
- Mitbestimmung stärken





KI ≠ KI

Ein Disclaimer



Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.



Angwin J, Larson J, Kirchner L, Mattu S (2016) Machine Bias - There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica
<https://www.rsna.org/news/2023/june/ai-improves-lung-cancer-detection>

Was ist KI?

Unterschiedliche Ansätze



Anfänge der KI

Eine Warnung

“ The science and engineering of making intelligent machines.

John McCarthy, 1956:

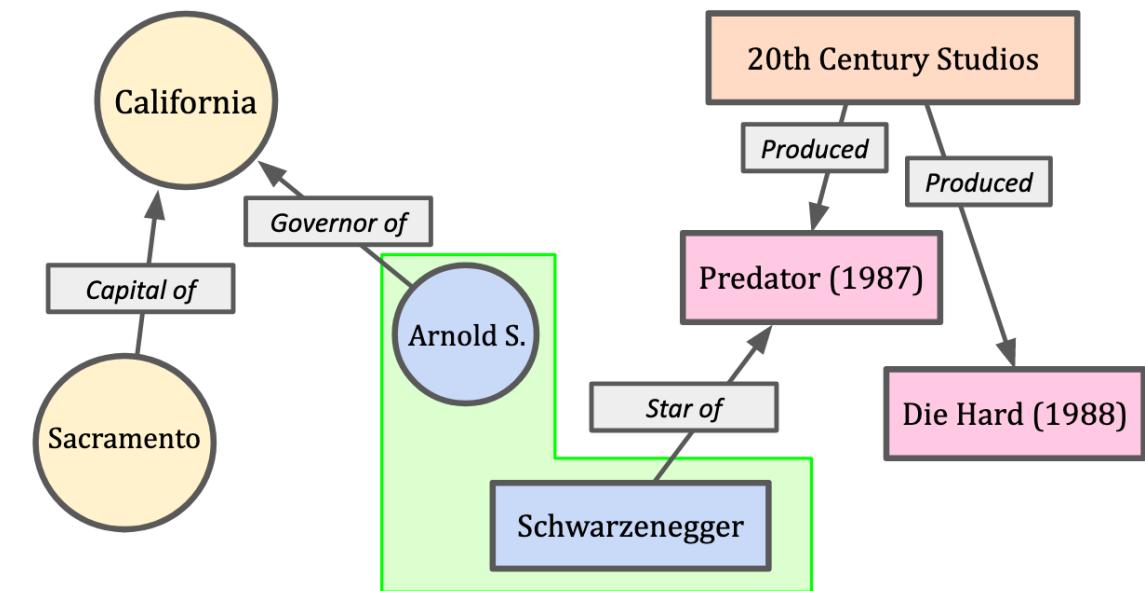
“ As soon as it works, no one calls it AI anymore.

- Artificial Narrow Intelligence (ANI) < Artificial General Intelligence (AGI) < Artificial Superintelligence (ASI)

Symbolische KI

Expert:innensysteme

- Explizite symbolische Repräsentation von Wissen, Logiken und Regeln — und damit auch zugrundeliegenden Annahmen.
- Versteht Wissen i. d. R. als idealerweise dekontextualisiert und entkörperlicht.

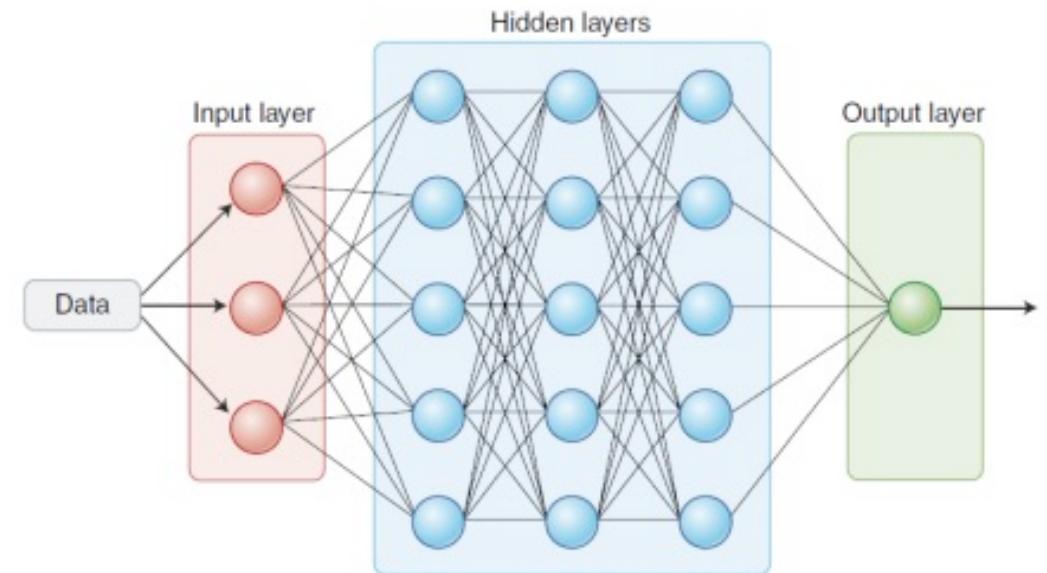


Subsymbolische KI

Konnektionismus

- Trainieren anhand größerer Datenmengen erlaubt induktive (nicht regelbasierte) und probabilistische (nicht kausale) Schlussfolgerungen.
- Mustererkennungen erfolgen anhand der statistischen Ähnlichkeit zu Trainingsdaten.
- Kernfrage:

“Wie wahrscheinlich ist eine Hypothese unter der Voraussetzung bestimmter Beobachtungen?”



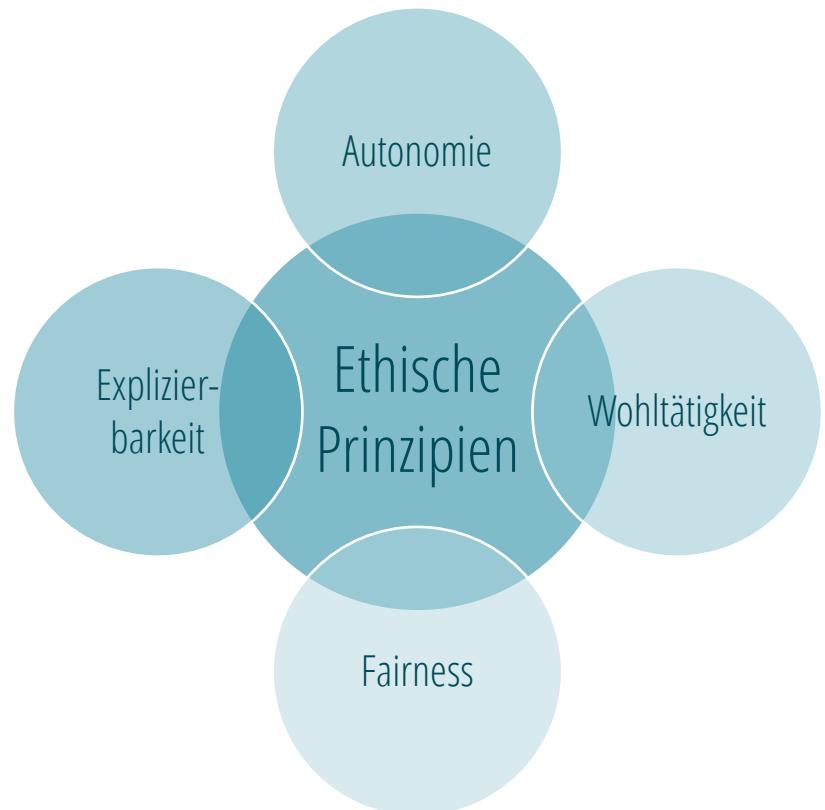
Wo stehen wir zur Zeit?

Aktueller Stand der ethischen Einhegung
Künstlicher Intelligenz



Vertrauenswürdige KI

Ethische Richtlinien der Europäischen Kommission

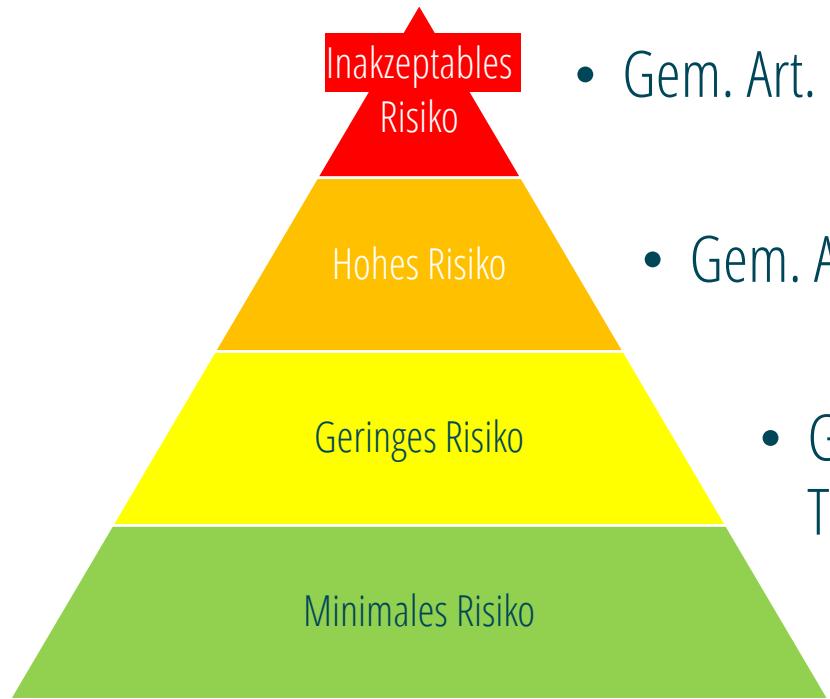


Independent High-Level Expert Group on Artificial Intelligence Set Up By the European Commission. (2019). Ethics Guidelines for Trustworthy AI.
<https://ec.europa.eu/futurum/en/ai-alliance-consultation>



Risikostufen

Wo liegen die ethischen Herausforderungen?



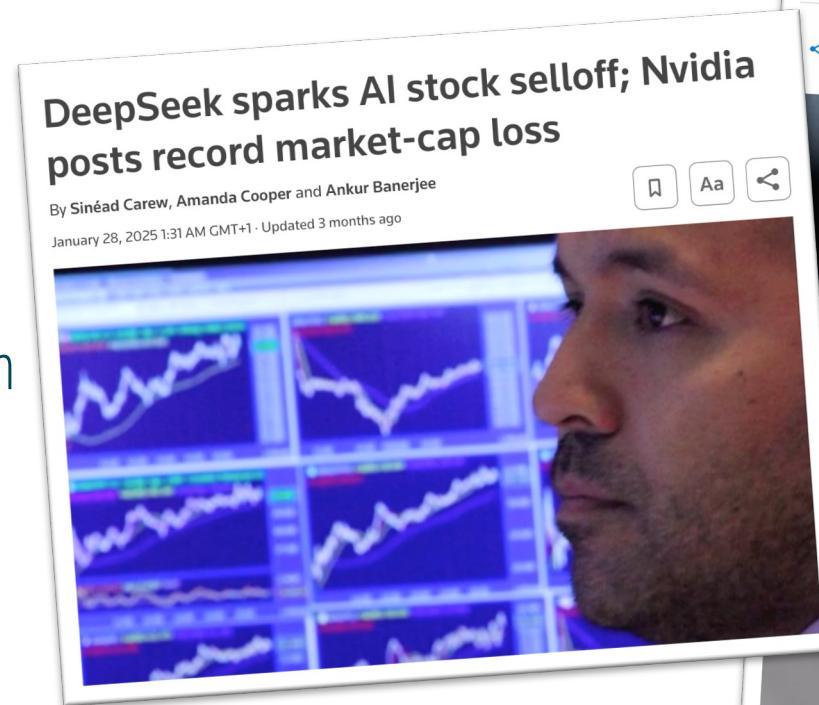
- Gem. Art. 5 KI-VO verboten
- Gem. Art. 6–51 KI-VO umfassend reguliert
- Gem. Art. 52 KI-VO gelten Transparenzpflichten
- Unterliegen keinen Beschränkungen



Vertrauenswürdige Innovation?

Hype-Kultur und volatile KI-Märkte

- Die Veröffentlichung des chinesischen DeepSeek (wesentlich effizienteres Computing) führte zur Devaluation der Nvidia-Aktie um 600 Mrd. \$ in nur einem Tag.

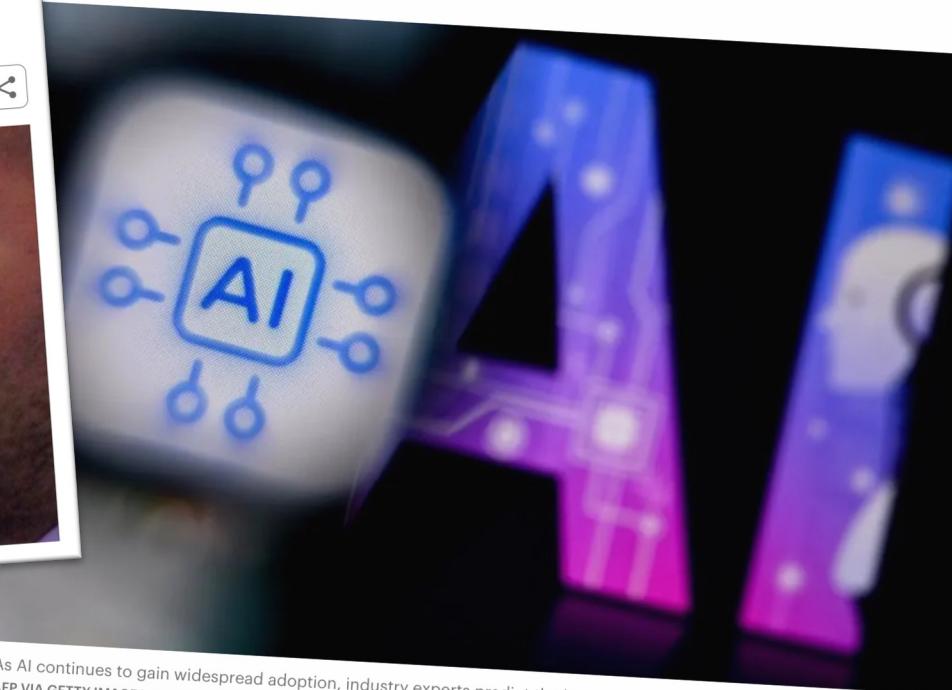


Experts Predict The Bubble May Burst For AI In 2025

By [Kolawole Samuel Adebayo](#), Contributor. ⓘ I write about the economics of AI.
Jan 20, 2025 at 12:33pm EST

Follow Author

Share Save Comment 1



Adebayo, K. S. (2025, January 20). Experts Predict The Bubble May Burst For AI In 2025. Forbes.
<https://www.forbes.com/sites/kolawolesamueladebayo/2025/01/20/experts-predict-the-bubble-may-burst-for-ai-in-2025/>

Mangelnde Explizierbarkeit

„Halluzinierende“ KI-Systeme

- KI-Systeme sind intransparenter und inexplizierbarer denn je.
- Generative KI „halluziniert“ mit steigender Tendenz zur Versionsnummer.
- Dies schließt auch Kenntnis zu gesellschaftlichen Auswirkungen ein.

Deming, D. J., Ong, C., & Summers, L. H. (2025). Technological Disruption in the Labor Market (Working Paper 33323). National Bureau of Economic Research. <http://www.nber.org/papers/w33323>

Anthropic CEO Admits We Have No Idea How AI Works

"This lack of understanding is essentially unprecedented in the history of technology."

MAY 4, 9:00 AM EDT by NOOR AL-SIBAI



Mangelnde Menschzentrierung

Wozu wird KI entwickelt?

- Trotz zahlreicher anderslautender Bekundungen trägt KI derzeit zu Jobverlusten bei. Die Gründe sind noch nicht vollständig verstanden.
- Gerade Einstiegs- und Kreativjobs scheinen gekürzt zu werden.
- Sogenanntes „AI Washing“ – KI als akzeptierte Begründung für Rationalisierung – scheint ebenfalls plausibel.

Thompson, D. (2025, April 30). Something Alarming Is Happening to the Job Market. The Atlantic. <https://www.theatlantic.com/economy/archive/2025/04/job-market-youth/682641/>



Haftungslücken

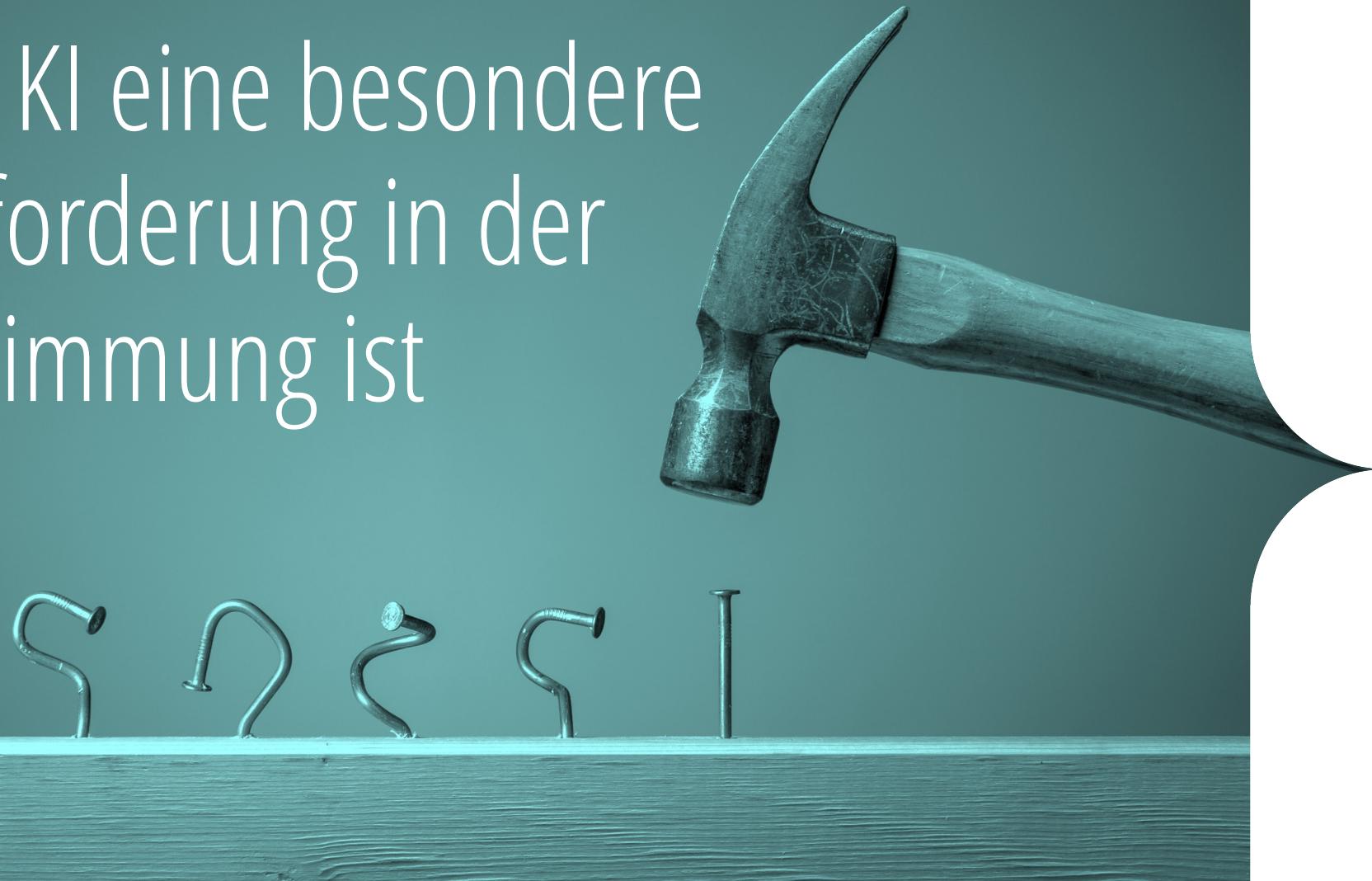
Generative KI

- Haftungsrahmen konzentrieren sich auf materielle Schäden und vernachlässigen immaterielle, finanzielle und gesellschaftliche Schäden durch Voreingenommenheiten, Halluzinationen und finanzielle Verluste aufgrund fehlerhafter KI-Produkte.



Wachter, Sandra. 2024. "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond." SSRN Electronic Journal, ahead of print. <https://doi.org/10.2139/ssrn.4924553>.

Warum KI eine besondere Herausforderung in der Mitbestimmung ist



These

Der Einsatz von Künstlicher Intelligenz in der Entscheidungsfindung fördert tendenziell Machtkonzentrationen.



KI & Macht

Möglichkeit der Zentralisierung

- Geschwindigkeit und Skalierbarkeit sind zentrale Machtfaktoren.
- Datenlieferant:innen sind “alle”, aber zur Auswertung und Entscheidung befähigt sind nur “wenige”.

KI als Chef

Was digitale Zwillinge von Führungskräften leisten

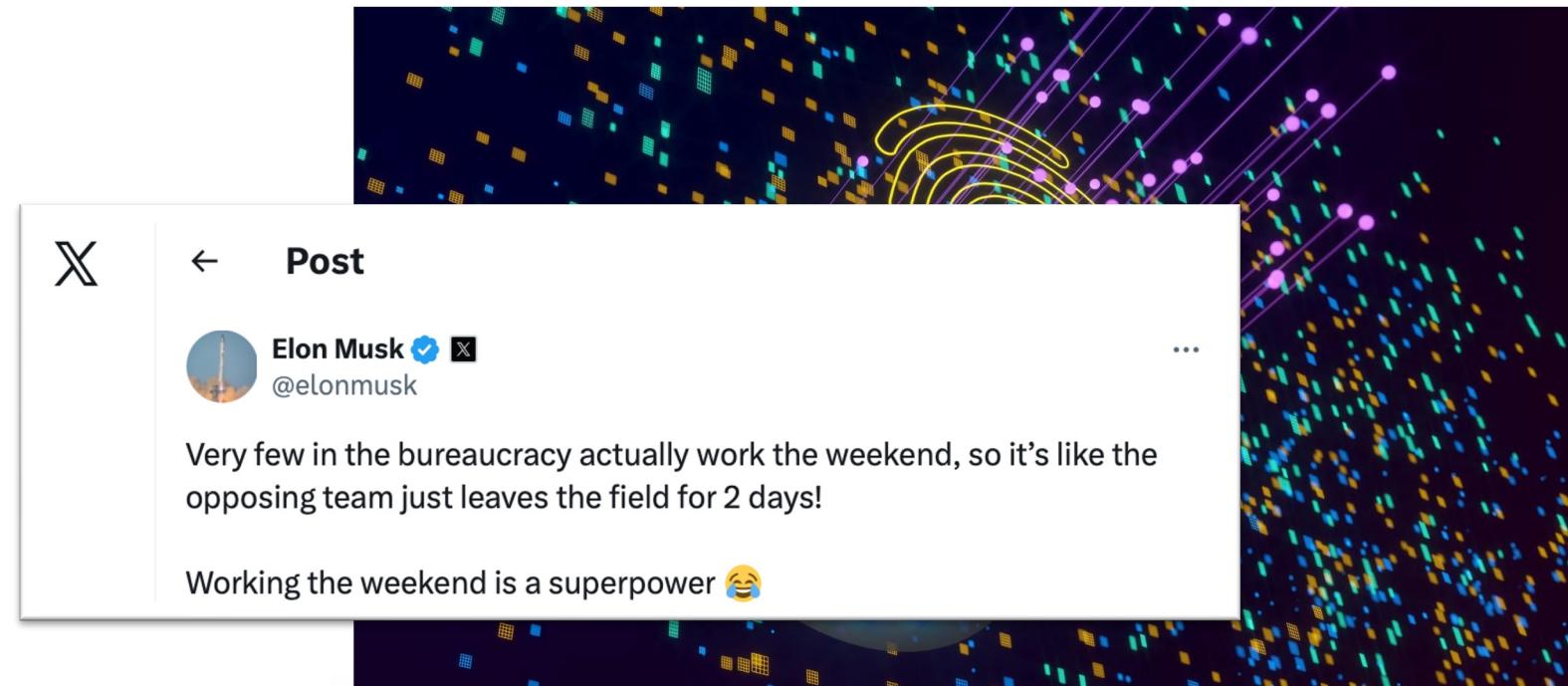
Expertenwissen bündeln und zugänglich machen, dafür nutzen viele Firmen Künstliche Intelligenz. Beim Logistikonzern Seifert testet man einen KI-Doppelgänger des CEO Axel Frey. Er soll sein Wissen für alle Mitarbeitenden verfügbar machen.

Krauter, Ralf; Schroeder Carina | 09. Oktober 2025, 09:00 Uhr

▶ Hören 33:02

⬇️ Audio herunterladen

Abonnieren  



X ← Post

Elon Musk ✅ X
@elonmusk

Very few in the bureaucracy actually work the weekend, so it's like the opposing team just leaves the field for 2 days!

Working the weekend is a superpower 😂

Soziotechnisches Phänomene

Eine subtil dominante Logik in der „wissenskulturellen Transformation“

- Nutzungskontexte und –interfaces unterliegen zunehmend der Logik statistischer Klassifikation.
 - Unser Leben im Digitalen wird in diese Logik eingehetzt, um die Daten nutzbar im Sinne der KI zu machen.
 - Durch die Nutzung derartiger Systeme, legen wir uns auch selbst diese Logik auf und verdrängen dadurch (auch im Kontext von Effizienzanreizen) Alternativen.

The future
works here.

Every candidate deserves to be heard. Eightfold AI Interviewer screens candidates and elevates top talent, so you can focus on final decisions, not first-round interviews.

See how it works

Datenherkunft, -qualität & -arbeit

Erfolgsfaktoren bei der Entwicklung datenbasierter Algorithmen

- Datenaufnahmen sind durch Anreizsysteme bedingt (Abrechnungsmodalitäten, etc.).
- Daten sind politisch: Datenerhebungen und –nutzungen greifen in die Möglichkeiten der Menschen ein, ihre Ziele zu erreichen. (Høyer, 2023)
 - Daten verkörpern Autorität über Wissen und Zugang zu diesem.
 - Daten werden genutzt, um Ressourcen zu alloklieren.
 - Daten verkörpern Werte und kommunizieren diese.
 - Daten konstruieren Populationen und Mittel der Beeinflussung.

Høyer, K. (2023). Data paradoxes: The politics of intensified data sourcing in contemporary healthcare. The MIT Press.

NAME: _____	DOB: _____
SEX: _____	DATE: _____
DR _____	

Patient Name:
Contact No.:
DOB:
Blood Group:
Medical Central

DATA PARADOXES

The Politics of Intensified
Data Sourcing in Contemporary
Healthcare

KLAUS HOEYER

NAME: _____
ID: _____ Age/Sex: _____
Weight: _____

Patient _____
DOB: _____ Age: _____ Sex: _____
Allergy: _____

Patient _____
Sex: _____ DOB: _____
General Hospital

Strukturelle Faktoren

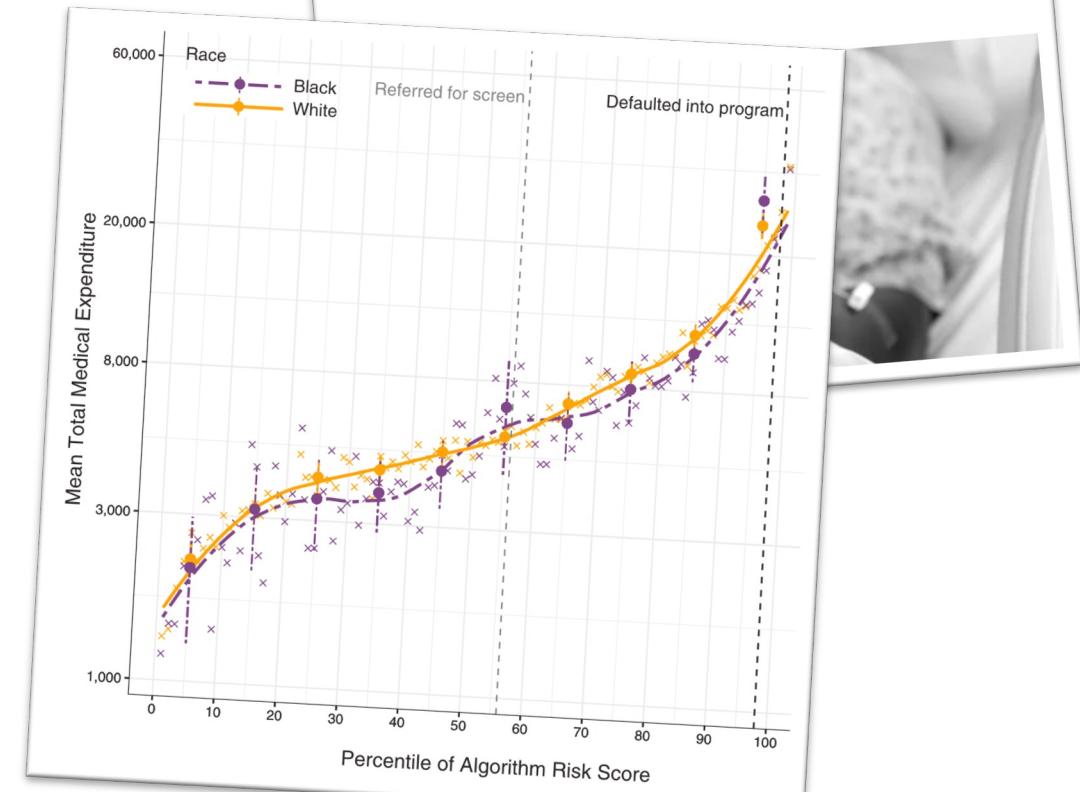
Beispiel für die strukturelle Perspektive: „Rassistische“ Algorithmen

- In einer Studie wurde festgestellt, dass ein häufig eingesetzter Algorithmus zur Bewertung von Gesundheitsrisiken rassistische Tendenzen aufweist.
- Die Bewertung des Bedarfs an medizinischer Versorgung wurde mit den historischen medizinischen Kosten als Surrogatparameter für die Prädiktion angenähert – eine Größe, die gesellschaftliche Rassismen annähert.

Racial Bias Found in a Major Health Care Risk Algorithm

Black patients lose out on critical care when systems equate health needs with costs

By Starre Vartan on October 24, 2019



Vartan, S. (October 24th, 2019). Racial Bias Found in a Major Health Care Risk Algorithm. Scientific American

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

KI-Basierte Entscheidungen

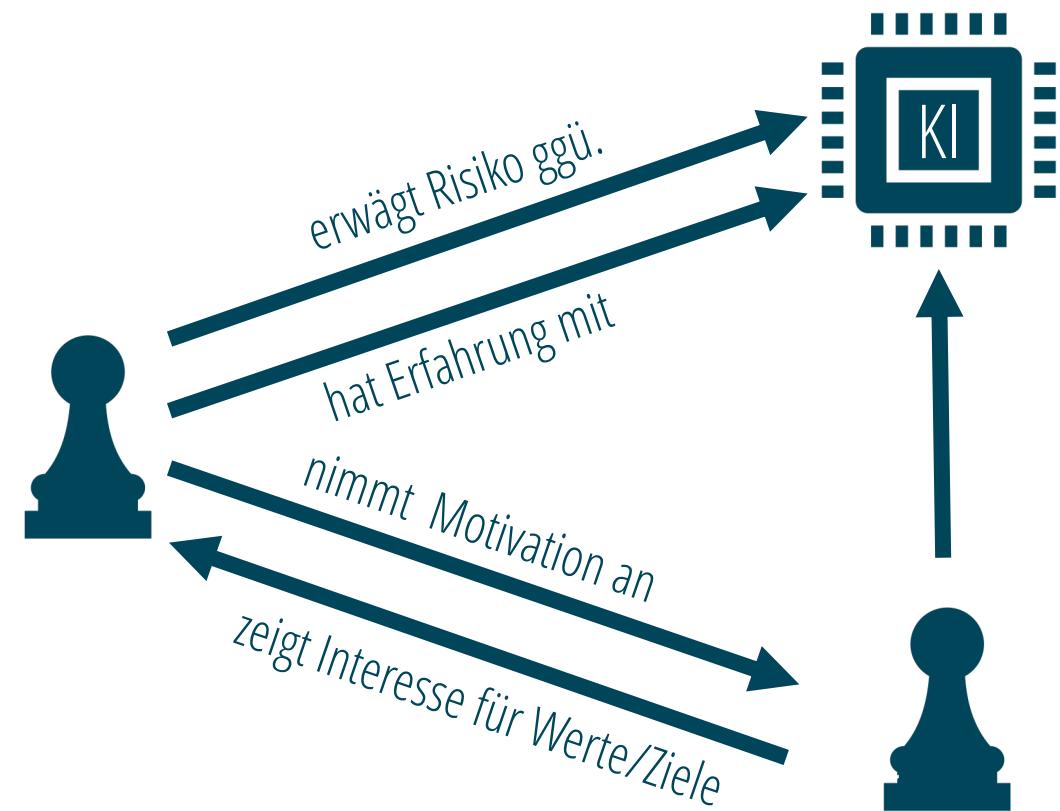
Ein Paradigmenwechsel



Implizite Werte in KI-Entscheidungen

Das Training einer KI ist selbst interessensgeleitet

- Grundsätzlich kann auf hohe Spezifität (geringe Falsch-Positive-Entscheidungen) oder hohe Sensitivität (geringe Falsch-Negativ-Entscheidungen) optimiert werden.
- Die Abwägung erfolgt bspw. danach, welche Fälle kritischer sind, folgt aber grundsätzlich dem Ansatz, eine große Anzahl von unterschiedlichen Fällen effizient zu bearbeiten.



KI in der Verwaltung

Zwei Paradigmen

- Regelbasierte Entscheidungen
 - Einzelfallentscheidungen sind aufwändig.
 - Informationsextraktion unterliegt im individuellen Fall einer Interpretationsleistung.
 - Die Aufstellung der Regeln unterliegen ggfs. moralischen/politischen Entscheidungen. Diese sind aber transparent.
- Konnektionistische Entscheidungen
 - Mimen Erfahrungen aus einer großen Anzahl an Fällen.
 - Sind bspws. darauf optimiert, dass statistisch gesehen „korrekt“ entschieden wird.
 - Bergen eine große Gefahr, insbesondere marginalisierte Gruppen falsch zu behandeln.

Toeslagenaffaire

Beispiel

- Algorithmisch vermittelte Aufdeckung von Sozialbetrug schadete Tausenden von Familien und blieb dennoch über sechs Jahre lang unentdeckt (Heikkilä, 2022).
- Es fehlte u.a. eine wirksame Benachrichtigung über die algorithmische Verarbeitung, inkl. Recht auf Widerspruch und die Forderung nach einer menschlichen Neubewertung, sowie das Recht auf eine Erklärung der Bedingungen und Parameter, die eine algorithmische Entscheidung beeinflusst haben.

FROM POLITICO PRO

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.



Heikkilä, M. (2022). Dutch scandal serves as a warning for Europe over risks of using algorithms. <https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/#>

Strukturelle Effekte

Unvereinbarkeit mit Rechtsstaatlichkeit

- Konnektionistische Entscheidungsfindung in der öffentlichen Verwaltung...
 - ...beruht auf Stereotypisierung und akzeptiert absichtlich Fehlentscheidungen im Einzelfall (vgl. Art. 3 GG),
 - ...gefährdet systematisch die Nutzung von Rechtsmitteln von Personen, deren Ressourcen begrenzt sind (vgl. Art. 19, Abs. 4, GG),
 - ...erhebt Effizienz zum maßgeblichen Prinzip,
 - ...untergräbt durch ihre Skalierbarkeit ggfs. das Prinzip der Kontrolle oder menschlichen Aufsicht,
 - ...erfährt einen selbstverstärkenden Effekt, wenn algorithmische Entscheidungen neue Daten für zukünftige Trainings generieren.

Mitbestimmung stärken

Künstlicher Intelligenz & Mitbestimmung?



These

Für die Stärkung guter Arbeit, müssen wir auf qualitative Zugewinne durch den Einsatz von KI setzen und den Wert von Arbeit und wo dieser entsteht reflektieren.



Beispiel: Generative KI

Potenziale?



Da virtuelle soziale Interaktionen zunehmend durch Algorithmen unterstützt werden, wird der Wert persönlicher Interaktionen, die von Maschinen nicht nachgebildet werden können, noch größer werden.



Aber wenn Sie etwas wirklich Neues schaffen wollen, anstatt bestehende Ideen wieder aufzuwärmen, worauf sollten Sie dann optimieren?

Frey, C. B., & Osborne, M. A. (2023, September 28). *Expert Comment: Jobs will be automated, but not because of the latest Generative AI*. University of Oxford News & Events. <https://www.ox.ac.uk/news/2023-09-28-expert-comment-jobs-will-be-automated-not-because-latest-generative-ai>

Expert Comment: Jobs will be automated, but not because of the latest Generative AI

ARTIFICIAL INTELLIGENCE EXPERT OPINION MATHEMATICAL, PHYSICAL AND LIFE SCIENCES RESEARCH SOCIAL SCIENCES

Everyone is worried about Artificial Intelligence. From writers in Hollywood to computer programmers, recent advances in technology are causing concern about what Generative AI is going to mean for the future of work, our society and the wider world. Is there nothing machines will not be able to do?

By [Professor Carl-Benedikt Frey](#), Dieter Schwarz Associate Professor of AI & Work, [Oxford Internet Institute](#) & Director, Future of Work Programme, Oxford Martin School, and

[Professor Michael Osborne](#), Professor of Machine Learning, [Department of Engineering Science](#) and co-Director, Oxford Martin AI Governance Initiative.

Was ist der Wert von Arbeit?

Interpretation und Adaption von Regeln und Prozessen

“

Regeln, auch technische, gelten nicht von selbst, sondern erfordern ein Subjekt, das sie in bestimmten Fällen anwendet.

Christophe Dejours

Dejours, Christophe (2018). *The Return of Work in Critical Theory: Self, Society, Politics*. New Directions in Critical Theory. New York: Columbia University Press.

The Return of
WORK in
Critical Theory

Self,
Society,
Politics



Christophe Dejours, Jean-Philippe Deranty,
Emmanuel Renault, and Nicholas H. Smith

Menschen und KI kollaborieren nicht

Menschliche Handlungsmacht darf nicht unterminiert werden

- Die Rede von der "Mensch-Al-Kollaboration" suggeriert auf problematische Weise eine Art von Handlungsgleichheit zwischen KI und Mensch.
- Vielmehr sind KI-Agenten nicht in der Lage, Zwecke zu verfolgen oder vielleicht sogar zu erkennen, die anders sind als jene, die im Entwurf vorgegeben sind.

Do We Collaborate With What We Design?

Katie D. Evans,^a Scott A. Robbins,^b Joanna J. Bryson^c

^a*Josef Stefan Institute (IRCAI), Sorbonne University*

^b*Center for Science and Thought, University of Bonn*

^c*Centre for Digital Governance, Hertie School*

Received 15 September 2022; received in revised form 20 June 2023; accepted 21 June 2023

Abstract

The use of terms like “collaboration” and “co-workers” to describe interactions between humans and certain artificial intelligence (AI) systems has gained significant traction in recent years. Yet, it remains an open question whether such anthropomorphic metaphors provide either even a purely innocuous lens through which to conceptualize designed commercial products or a potentially misleading one.

Ansatz

Projekt gefördert von der Hans-Böckler Stiftung

Hans **Böckler**
Stiftung ■ ■

- Stärkung des konstruktiven Potenzials in der Mitbestimmung (Krzywdzinski 2024)
- Ausarbeitung: Workshopkonzept + Canvas
- Challenge-Based Learning (Nichols und Cator 2008)
- Iterativ, partizipativ und modular



**STADTWERKE
LÜBECK**



Nichols, Mark H., and Karen Cator. 2008. Challenge Based Learning White Paper. Apple.



Situation der Betriebsräte

Herausforderungen

- Rasante technologische Entwicklung
- Hochgradig diverse fachliche Hintergründe und Ausbildungsniveaus
- Oftmals berufliche Doppelbelastung



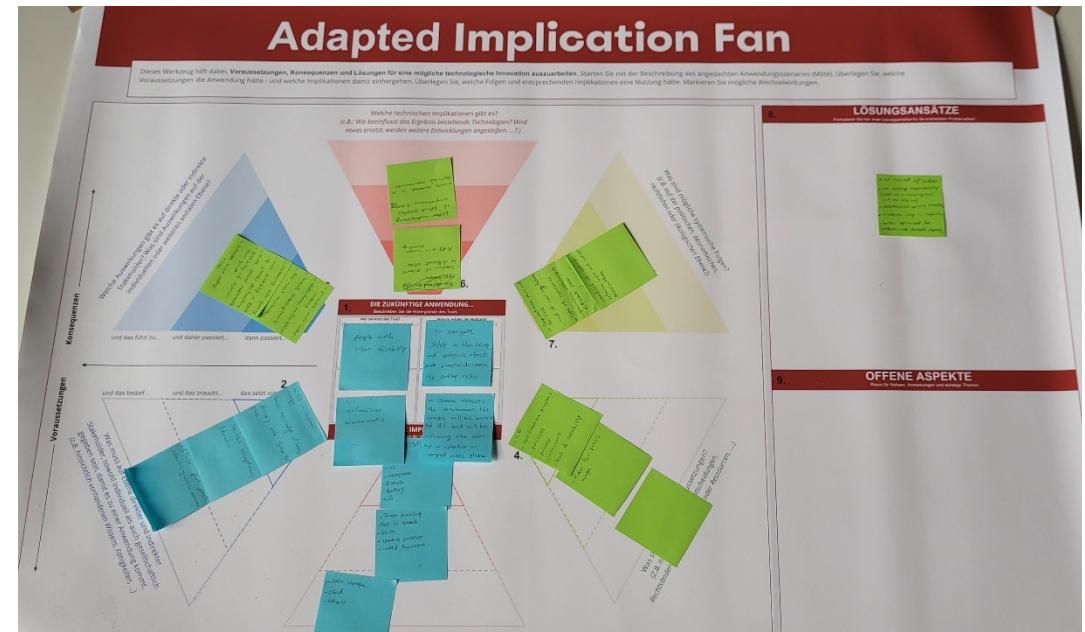
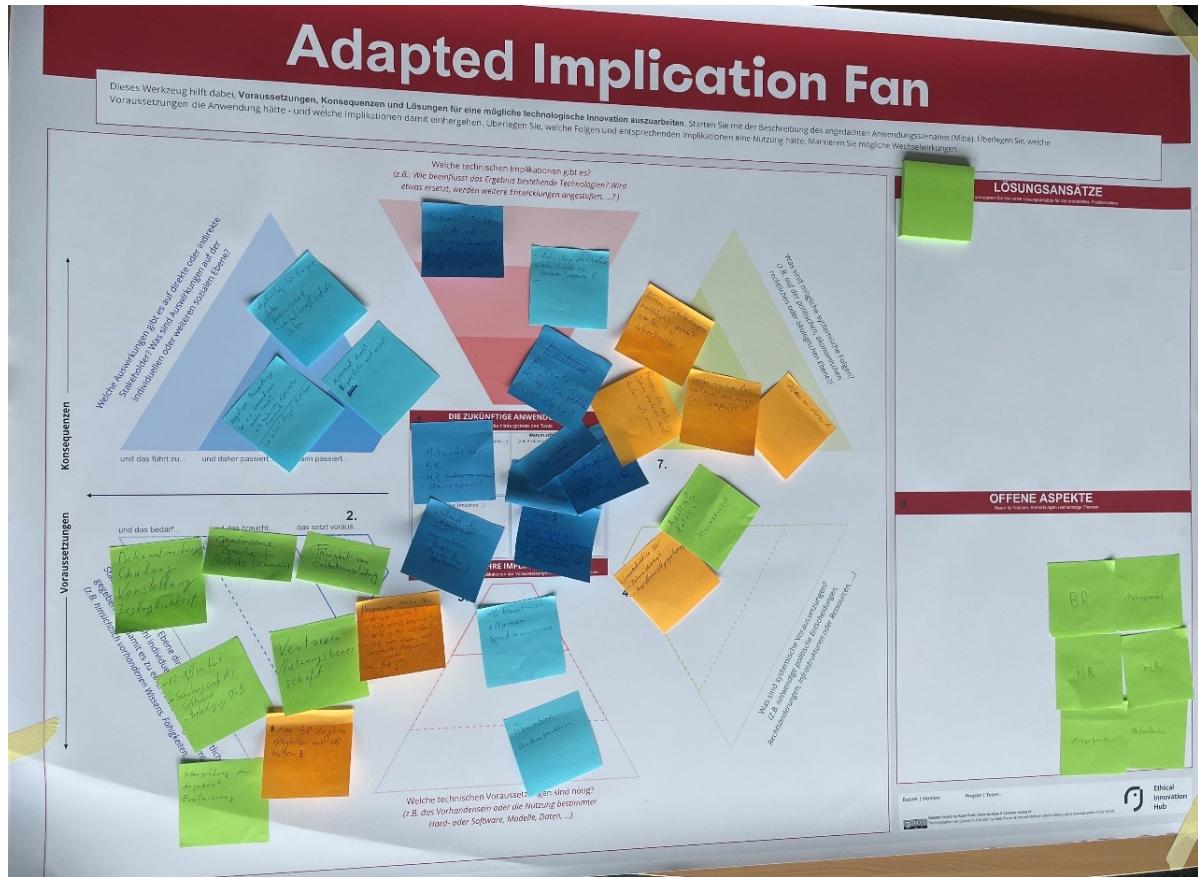
Ziele

Projektarbeit für mehr Partizipation in der Einführung betrieblicher KI

- Methodische und kritisch-reflexive Kompetenz der Betriebsräte für eine konstruktive, argumentationsbasierte Mitbestimmung stärken
- Sozioethische Unterstützung und Bewusstseinsbildung
- Sensibilisierung für KI-Risiken aber auch Chancen
- Nivellierung epistemisches Ungleichgewicht



Methodisches Vorgehen



Resümee und Ausblick

- KI zentralisiert tendenziell Macht.
- In der Gestaltung der digitalen Transformation muss auf stärker auf qualitative Zugewinne geachtet werden.
- Es bedarf einer Stärkung der betrieblichen Mitbestimmung, u.a. über den expliziten und methodischen Einbezug ethischer Aspekte.





Ethical
Innovation
Hub



UNIVERSITÄT ZU LÜBECK

Vielen Dank



Prof. Dr.-Ing. Christian Herzog, M.A.
Leiter des Ethical Innovation Hub

+49 451 3101 6211
christian.herzog@uni-luebeck.de
www.eih.uni-luebeck.de

