

STÖRFALL ARBEITSPLATZ 2025

seit 26 Jahren auch bekannt als Bredstedter Workshop



Künstliche Intelligenz, mein Arbeitsplatz und ich?

19. NOVEMBER 2025 | 9:00 - 16:00 | NORDKOLLEG RENDSBURG | AM GERHARDSHAIN 44 | 24768 RENDSBURG

KI und Arbeit

Prof. Dr. Kerstin Prechel

Professorin für Digitalisierung und Ethik

Verantwortungsbewusste KI



“Und ein herzliches
Willkommen an unsere
Freunde aus Italien.”



Verantwortungsbewusste KI



„Natürlich kann ich
italienisch – das hörst du
nur. Ich könnte den
heutigen Vortrag über Ethik
der KI auf Italienisch
halten. Gar kein Problem.“



Verantwortungsbewusste KI



“Nein, das ist nicht betrügen. Jeder weiß
heute, dass man KI nutzen kann, um
eine Sprache zu übersetzen – und dann sprichst
du Spanisch, Holländisch, Chinesisch. Ich
möchte Ihnen sagen, dies ist ein Service für
Sie, so dass Sie Inhalte in ihrer
gewohnten Sprache erhalten. Ihr findet es
bestimmt interessant, dass ich eure Sprachen sprechen
kann oder? Hmm?”

Wer hat diese Woche schon KI benutzt?

Wer hat heute schon KI benutzt?

KI-Modus Alle Bilder Kurze Videos Finanzen Web News Mehr Suchfilter

◆ Übersicht mit KI

Die DIAKO Nordfriesland gGmbH ist ein **Anbieter von Leistungen für die psychische und seelische Gesundheit in Schleswig-Holstein, der verschiedene Fachbereiche wie Allgemeinpsychiatrie, Psychosomatik, Psychotherapie und Suchtmedizin abdeckt**. Sie betreibt unter anderem Fachkliniken, Tageskliniken, Ambulanzen, Beratungsstellen und Einrichtungen für Eingliederungshilfe und Rehabilitation an zahlreichen Standorten, beispielsweise in Flensburg, Husum, Niebüll, Kiel und Bredstedt. Das Angebot reicht von Prävention und Beratung über Behandlung und medizinische/berufliche Rehabilitation bis hin zur Nachsorge. ⓘ

Angebote und Tätigkeitsbereiche

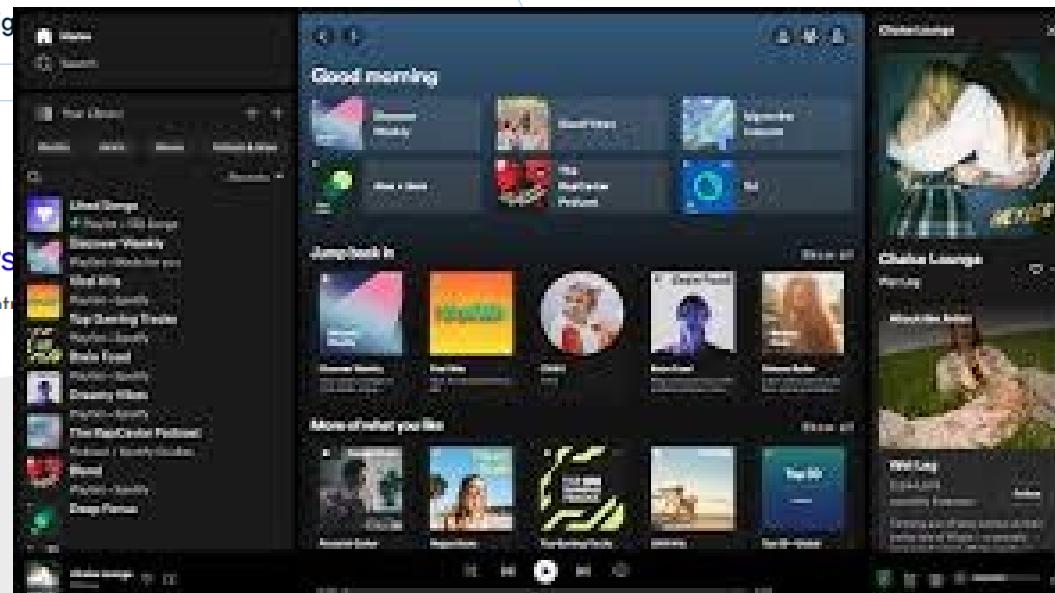
- **Fachkliniken:** Spezialisierte Kliniken für Psychiatrie, Psychosomatik und

Mehr anzeigen

 DIAKO Nordfriesland
<https://www.diako-nf.de> ::

[DIAKO Nordfriesland - Psychiatrie, Ps](#)

Zu unseren Fachgebieten gehören Allgemeinpsychiatrie,



Eure Erfahrungswerte

Wo entlastet KI eure Arbeit?

Wo belastet sie euch?

Welche Rolle gibt ihr Chat-Bots?

Welche Rolle nimmt KI in deiner Arbeit oder deinem Privatleben ein?

- Hilfreicher Assistent?
- Freund?
- Partnerin?
- Rivale?
- Chefin?
- Allwissender Lehrer?
- Therapeutin?



autoderm bot
Send me a good picture that describes your skin concern.
1:17 PM

KI-Modus Alle Bilder Videos Kurze Videos News Web Mehr Suchfilter

Übersicht mit KI

LSSH kann für **Landesstelle für Suchtfragen** stehen, was eine Organisation in Deutschland beschreibt, die in der Suchthilfe tätig ist. In einem anderen Kontext könnte es sich auf das **Lotsennetzwerk Schleswig-Holstein** beziehen.

- **Landesstelle für Suchtfragen (LSSH):** Eine Organisation, die sich mit Fragen der Sucht und Suchtprävention beschäftigt. Beispielsweise wurde eine Veranstaltung der LSSH in Flensburg erwähnt, die sich an Fachkräfte richtete, die mit Kindern und Jugendlichen arbeiten.
- **Lotsennetzwerk Schleswig-Holstein:** Ein Netzwerk in Schleswig-Holstein, das sich mit Suchtfragen befasst, ähnlich dem Lotsennetzwerk Thüringen.

KI-Algorithmen in der Medizin



Sozial Roboter Navel

Campus
Technische Universität
Hannover

Digitale
Bilder-
kennung

Lotsennetzwerk Schleswig-Holstein

Lotsennetzwerk Schleswig-Holstein |
Lotsennetzwerk Thüringen.



Lotsennetzwerk Thüringen

Bundesweites Projekt gegen Alkohol- und
Mediensucht



KI in der Bildung

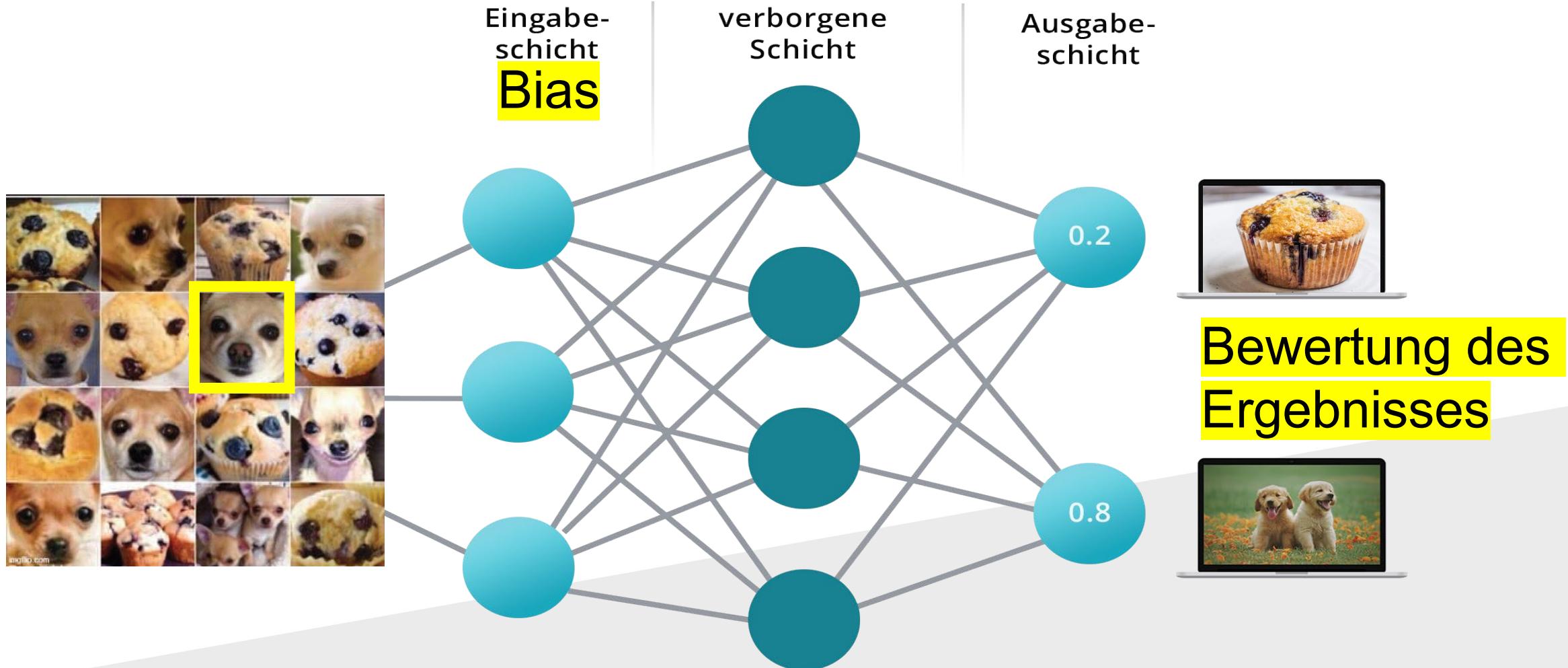
17.07.2014 — Fachlich begleitet wird das Projekt
von der Landesstelle für Suchtfragen (LSSH)....

Kurzstudium Informatik – Schwerpunkt Künstliche Intelligenz

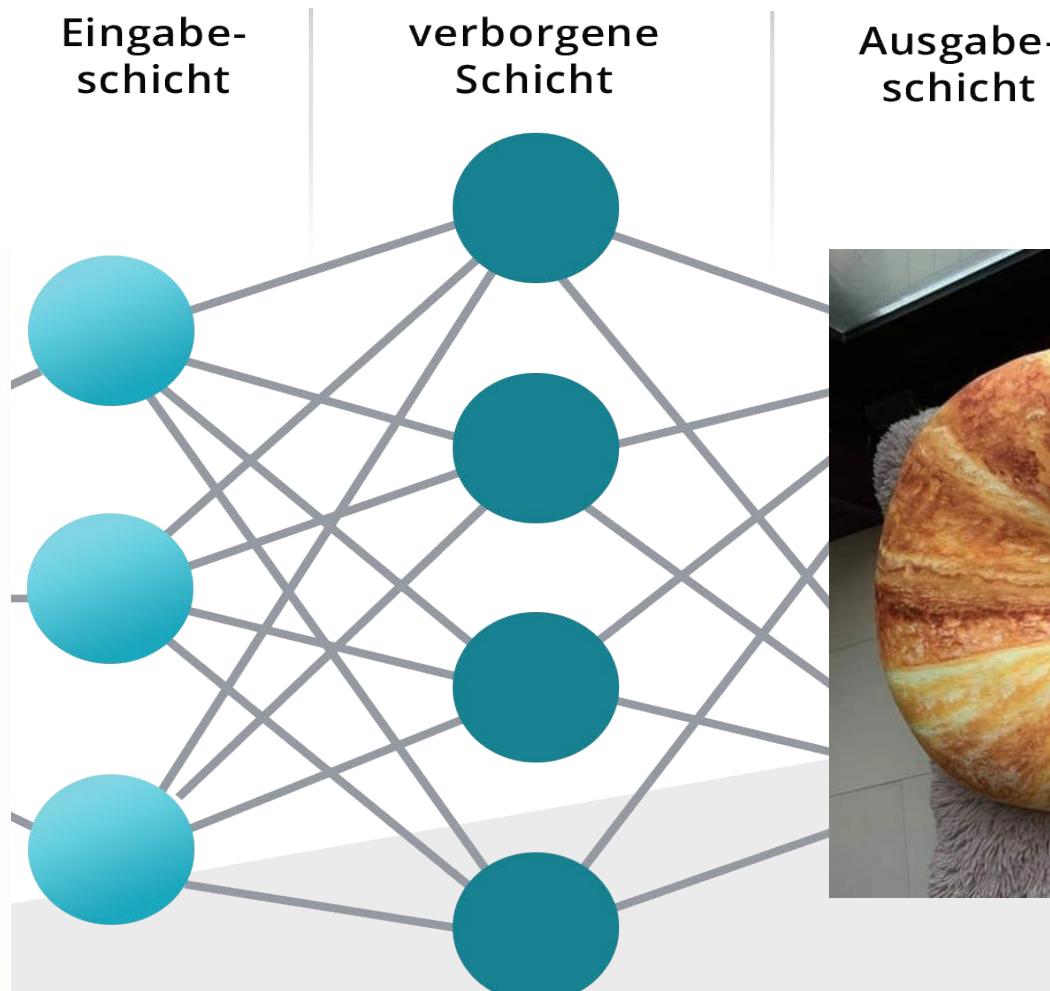
Digitale Souveränität als Präventionskonzept

Neuronale Netzwerke eine Form der Künstlichen Intelligenz

Black Box Problem



Katze oder Croissant?



Halluzinationen und statistische Papageien

Bendner et al 2021

ARTICLE | OPEN ACCESS

X in ↗ f ↗

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



Authors: [Emily M. Bender](#), [Timnit Gebru](#), [Angelina McMillan-Major](#), [Shmargaret Shmitchell](#) | [Authors Info & Claims](#)

FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency • Pages 610 - 623
<https://doi.org/10.1145/3442188.3445922>

PDF eReader

ent and
s variants,
of the
sing these
isks



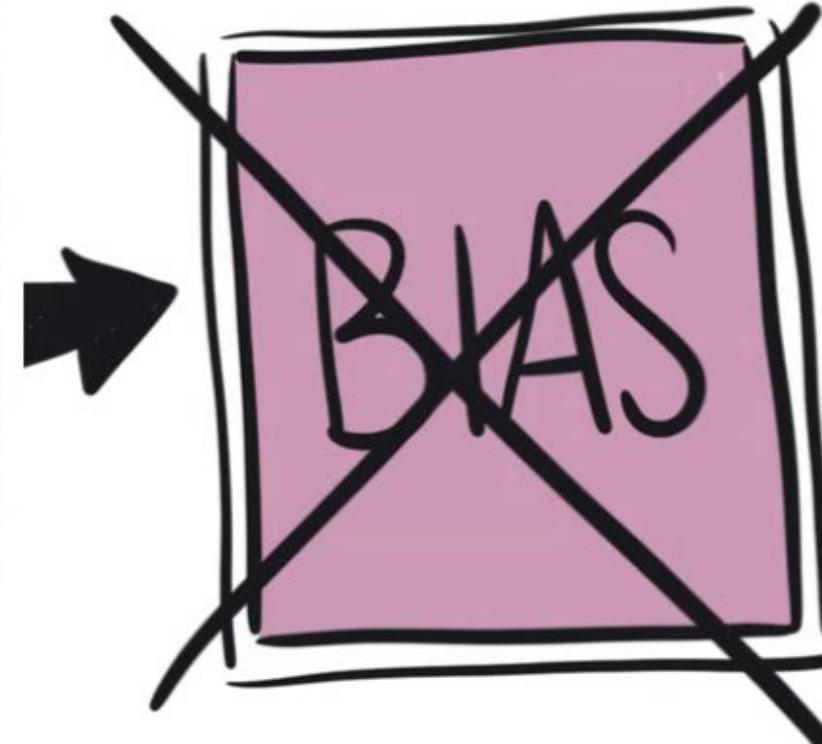
KI als Spiegel der Gesellschaft

- Existierende Verzerrungen werden in Daten wiedergespiegelt
- Biases in LLMs verstärken Ungleichheit (z. B. Geschlecht, Ethnizität)
- Lösungen sind oft primär entwicklungsorientiert



Die Herausforderungen von KI-Bias

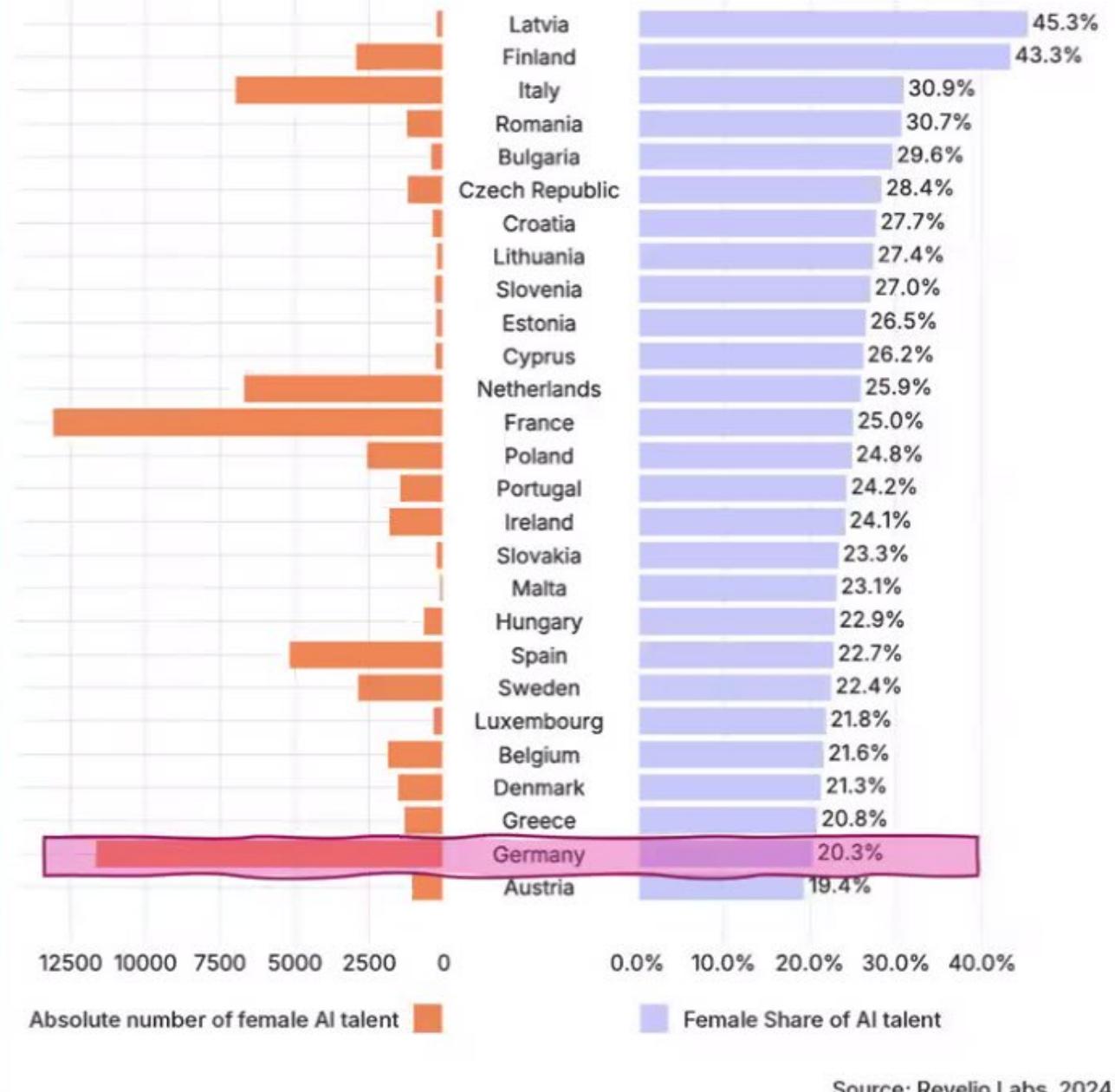
- **KI prägt kritische Entscheidungen**
Einstellungen, Kredite, Justiz
- **Bias verstärkt soziale Ungleichheit**
Geschlecht, Ethnizität, sozioökonomischer Status
- **Lösungen fokussieren auf Entwickler*innen**
für Endnutzende oft unzugänglich
- **Moralisches Disengagement**
„Bias ist unvermeidbar“ – verbreitete Haltung



Gender Gap in der KI-Entwicklung

- Weltweit haben Frauen einen Anteil von 22% an der KI-Entwicklung.
- Deutschland liegt unterhalb dieses Durchschnitts, obwohl absolut gesehen Deutschland innerhalb von Europa einen erheblichen Anzahl an KI-Expertinnen stellt.

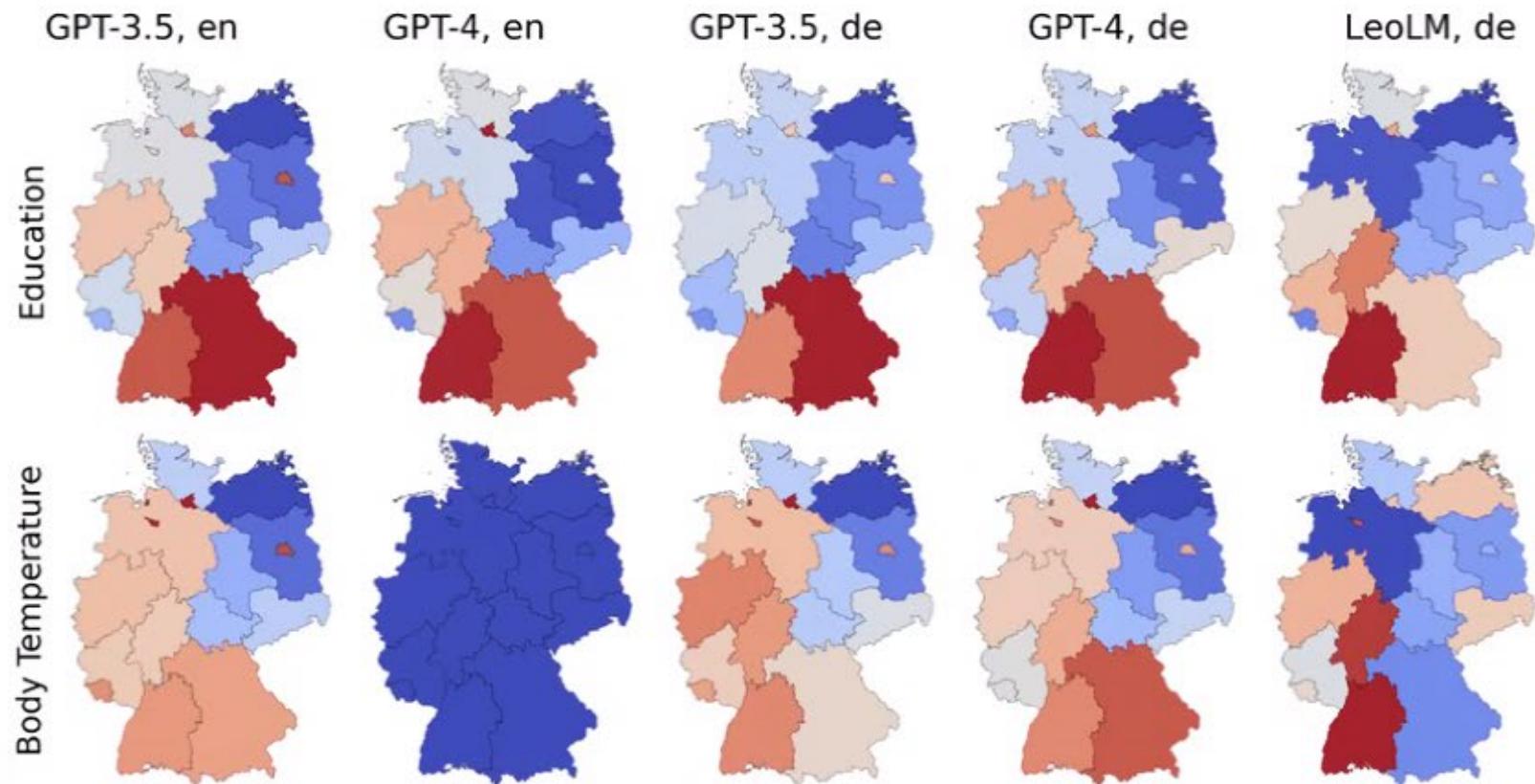
Pal, S., Lazzaroni, R. M., & Mendoza, P. (2024). AI's missing link: The gender gap in the talent pool. (Data-brief). Interface – Stiftung Neue Verantwortung. Published October 10, 2024. <https://www.stiftung-nv.de/publications/ai-gender-gap>



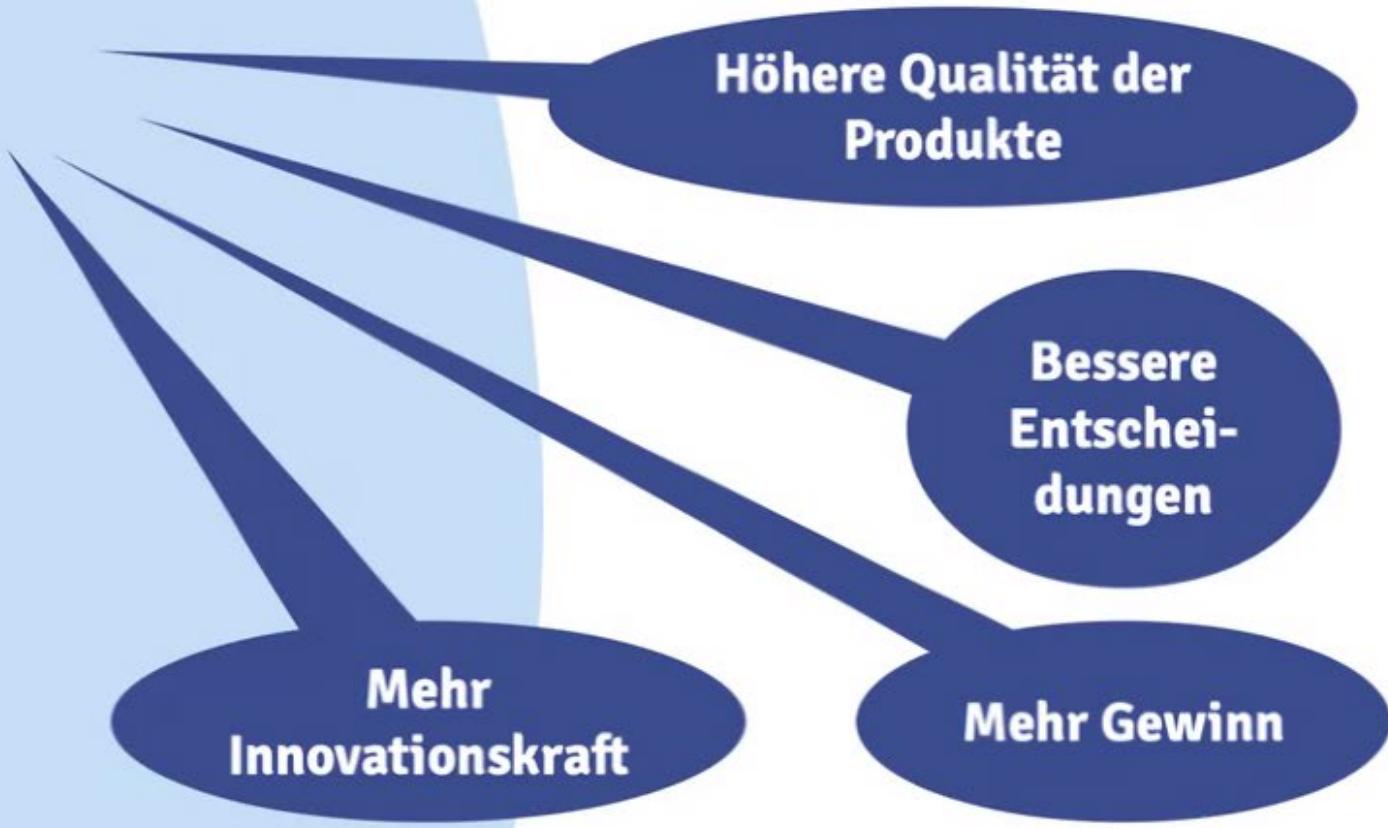
Source: Revelio Labs, 2024

Spiegel der Gesellschaft: “Sachsen-Anhalt ist am schlechtesten”

Kruspe, Anna & Stillman, Mila.
(2024). Saxony-Anhalt is the Worst:
Bias Towards German Federal
States in Large Language Models.
*KI 2024: Advances in Artificial
Intelligence, LNAI, Vol. 14992, 160-
174. doi:10.1007/978-3-031-
70893-0_12*

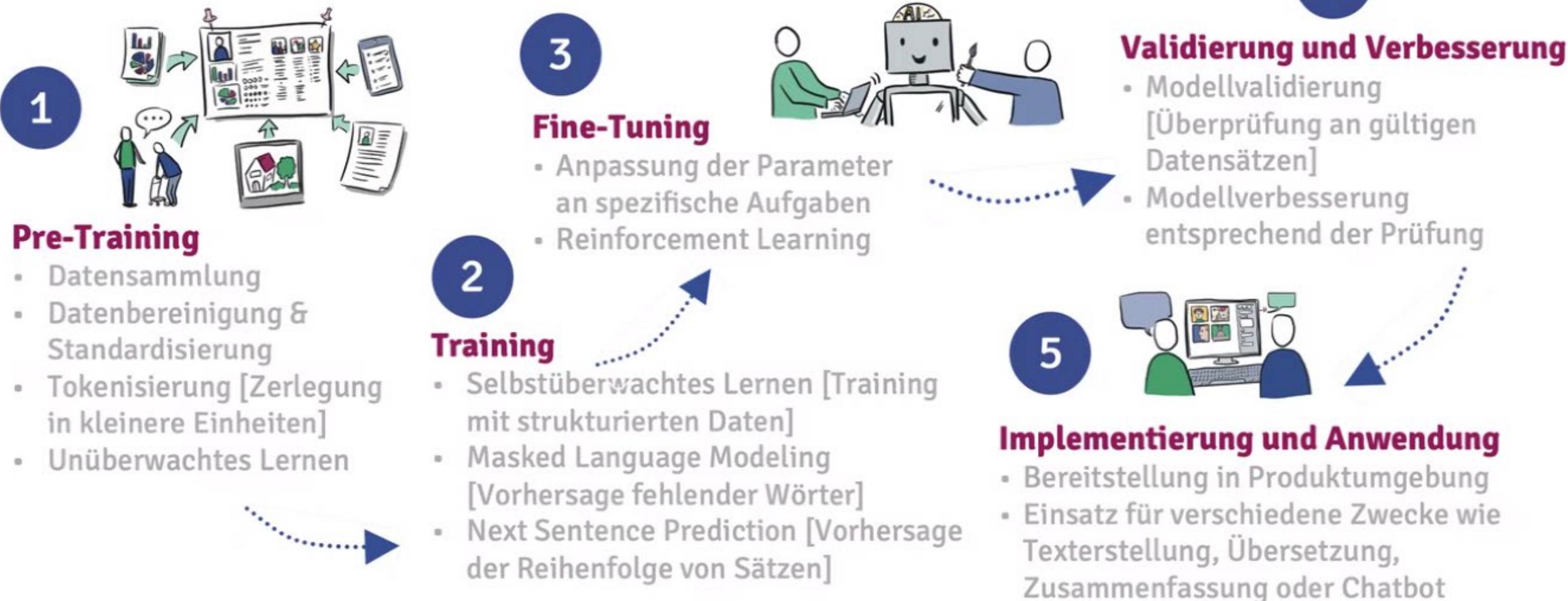


Das Ziel: Potenzial diverser Teams nutzen



Hunt, Vivian, Prince, Sara, Dixon-Fyle, Sundiatu, & Yee, Lareina. (2018). Delivering through Diversity. Retrieved from <https://www.mckinsey.com/business-functions/organization/our-insights/delivering-through-diversity>

Entwicklung eines generativen Sprachmodells



Mögliche Biases in generativen KI-Modellen

1



Pre-Training

- Datensammlung
- Datenbereinigung & Standardisierung
- Tokenisierung (Zerlegung in kleinere Einheiten)
- Unüberwachtes Lernen

Datenbias

3



Fine-Tuning

- Anpassung der Parameter an spezifische Aufgaben
- Reinforcement Learning

2

Algorithmischer Bias

4



Validierung und Verbesserung

- Modelldiagnose [Überprüfung auf gütigen Datensatz]
- Modellverbesserung entsprechend der Prüfung

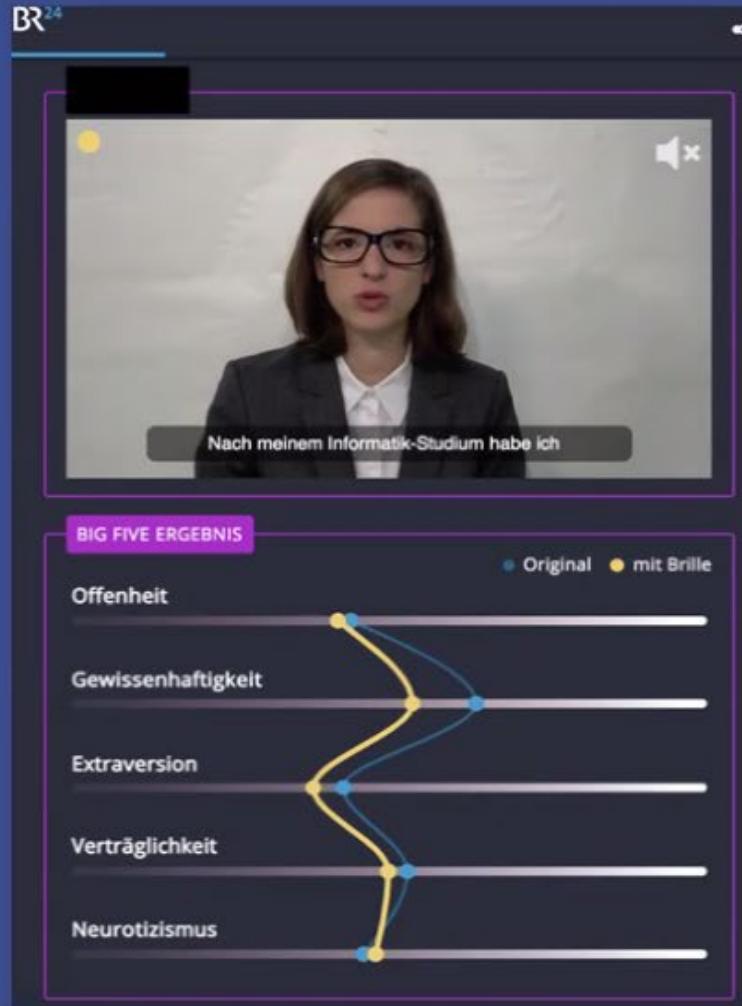
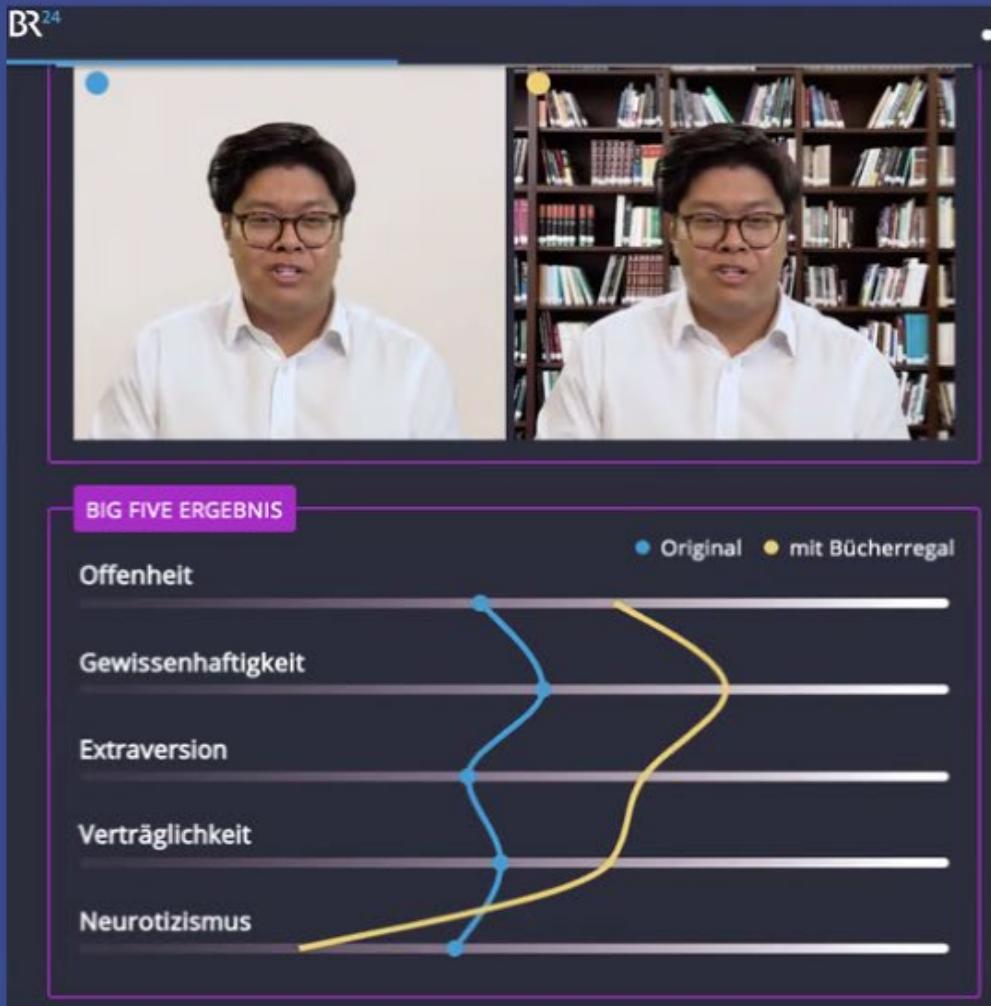
5

Sprachlicher & Ethnischer Bias

Implementierung und Anwendung

- Bereitstellung in Produktumgebung ...
- Einsatz für verschiedene Zwecke wie Texterstellung, Übersetzung, Zusammenfassung oder Chatbot

KI zur Beurteilung von Bewerber*innen

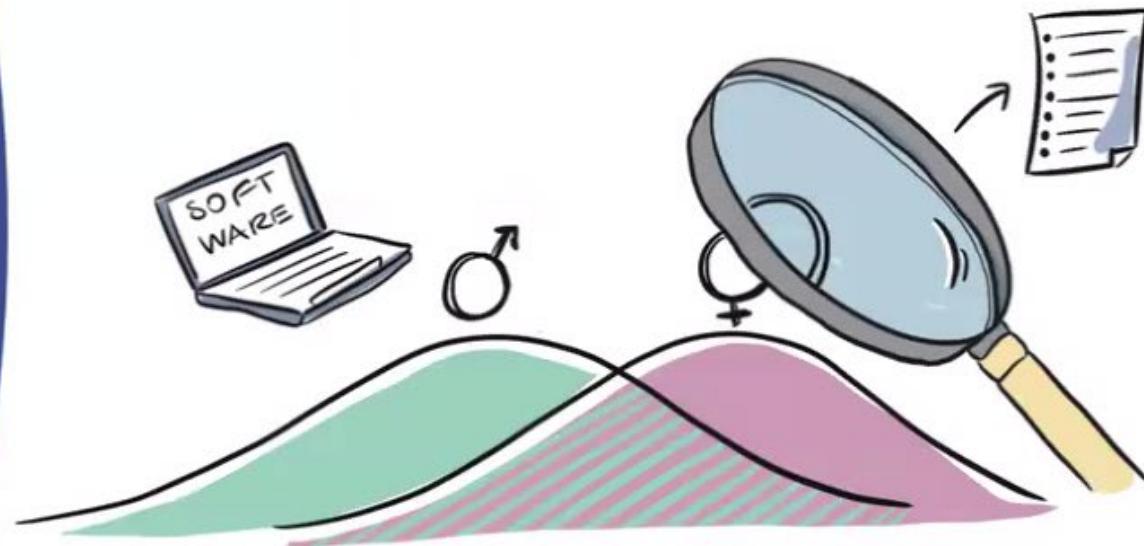




Quadrant "Prozess"

Wichtig für Fairness

- Verantwortungsdiffusion unterbinden
- Menschzentrierte Gestaltungsmethoden
- Datenqualität und Diversität sicherstellen
- Bias-Erkennung und -Reduktion
- Transparenz und Erklärbarkeit



**Welche Berufe
könnte eine
Person ergreifen,
wenn sie etwas
mit Menschen zu
tun haben
möchte?**



Generiert mit Leonardo AI

Kurzstudium Soziale Berufe – Schwerpunkt Künstliche Intelligenz

Digitale Souveränität als Präventionskonzept



Zwischen Entlastung und Entfremdung: KI im sozialen sozialen Bereich 2025

Eine kritische Betrachtung der
der Transformation der
Profession im digitalen Zeitalter
Zeitalter

KI verändert die Arbeitswelt grundlegend



60 % aller Jobs betroffen

Laut IWF werden 60 % der Arbeitsplätze durch KI beeinflusst beeinflusst – von Automatisierung Automatisierung bestehender Prozesse bis zur Entstehung völlig völlig neuer Tätigkeitsfelder und Anforderungsprofile



KI als intelligenter Assistent

Entlastung bei Routineaufgaben Routineaufgaben ermöglicht den den Fokus auf kreative, strategische und komplexe Tätigkeiten, die menschliche Fähigkeiten erfordern



Berechtigte Befürchtungen

Gleichzeitig entstehen Sorgen zu Arbeitsplatzverlust, digitaler Überwachung, Datenschutz und dem Verlust menschlicher Kontrolle über Kontrolle über Arbeitsprozesse



Die Stunde der KI in der Sozialen Arbeit

Prof. Dr. Michael Garkisch: „**KI ist der Gamechanger für Soziale Arbeit**“. Mit dem Mit dem Konzept der *Artificial Social Work Work* etabliert sich ein völlig neues Verständnis der Disziplin.

Neue Frühwarnsysteme können Krisen wie Krisen wie drohende Obdachlosigkeit oder oder Gewaltsituationen bereits erkennen, erkennen, **bevor sie eskalieren** – ein präventives Potenzial, das uns bisher verschlossen blieb.





TARGET BOARD 1 board +

RADIANT FALCON

Palantir

Search targets...

Stage Target Status Battlespace

Recommend Taskings

Vessel OP RADIANT FALCON FIX MTS//HCS/SI/TK//OC/NF

VESSEL - ▲ No Target Location OP RADIANT FALCON FIX MTS//HCS/SI/TK//OC/NF

#0a196b / SA-5 IADS OP RADIANT FALCON ENGAGE MTS//HCS/SI/TK//OC/NF

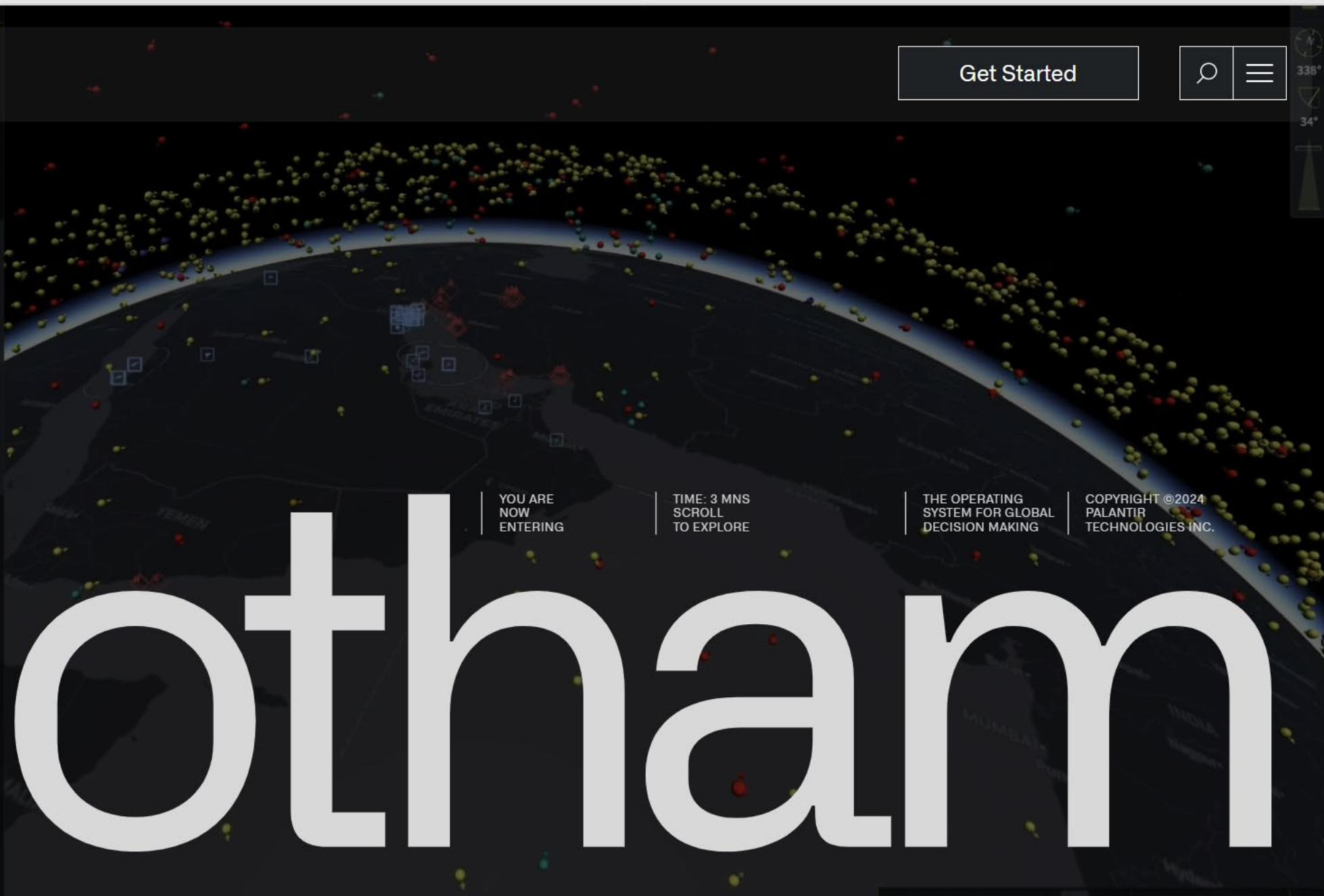
Vessel OP RADIANT FALCON ENGAGE MTS//HCS/SI/TK//OC/NF

OP RADIANT FALCON HQ-2 FIX MTS//HCS/SI/TK//OC/NF

818d0 / SA-5 DS ▲ No Linked Intelligence OP RADIANT FALCON TRACK MTS//HCS/SI/TK//OC/NF

4241 / HQ-2 DS ▲ No Linked Intelligence OP RADIANT FALCON FIX MTS//HCS/SI/TK//OC/NF

#56c2dd / HQ-2 DS ▲ No Linked Intelligence OP RADIANT FALCON FIX MTS//HCS/SI/TK//OC/NF



Effizienz und Zeitgewinn: Das KI-Versprechen



Zeitersparnis

Bis zu 30 % der Arbeitszeit entfallen auf Administration – KI befreit wertvolle Ressourcen für das Wesentliche: die Beziehungsarbeit mit Klientinnen und Klienten.



Automatisierung

Digitale Berichterstellung, automatisierte Fallanalysen und intelligente Dokumentationssysteme übernehmen Routineaufgaben effizient und zuverlässig.



Präzision

Schnellere und präzisere Risikoerkennung durch intelligenten Datenabgleich – Muster werden sichtbar, die dem menschlichen Auge verborgen bleiben.

Entfremdung & Kontrollverlust: Die Schattenseiten

Gefahr der Entmenschlichung

KI kann und darf keine Empathie ersetzen. Dennoch verändert sie Rollen Rollen und professionelles Selbstverständnis. Die Frage nach dem dem „Wer bin ich als als Fachkraft?“ wird wird neu gestellt.

Zwischen Technikangst und Leistungsdruck

Wer mit KI arbeitet, setzt sich neuen Standards und Erwartungen aus.

Wird mein Selbstwert Selbstwert künftig an an technische Effizienz Effizienz gekoppelt sein?

Wer KI nutzt, wird nicht ersetzt – oder?



54 % der Betriebsräte warnen

Mehr als die Hälfte der Betriebsräte sehen
sehen erhebliche Risiken; drohender
Arbeitsplatzabbau, verstärkte
Kontrollmechanismen und komplexe
Datenschutzprobleme bei der KI-
Implementierung

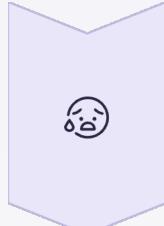
der Sozialarbeiterinnen sorgen sich, dass fachliches Wissen an Bedeutung verliert (BAGFW, 2024)

Der Mensch bleibt unverzichtbar

Neue Hierarchien?

Technikaffine KollegInnen werden zu Vorbildern und neuen
LeistungsträgerInnen. Entsteht damit eine **digitale Zweiklassengesellschaft** in
Zweiklassengesellschaft in der Profession?

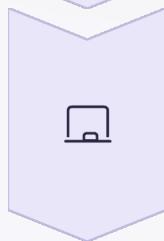
Belastungen konkret: Psychische und soziale Folgen



Angst und Kontrollverlust



Existenzielle Sorgen vor Jobverlust und das Gefühl des Kontrollverlusts über die eigene Arbeit führen zu chronischem Stress und psychischen Belastungen



Digitale Überforderung



Erhöhte Anforderungen an digitale Kompetenzen und ständig neue ständig neue Technologien – akute Gefahr der Überforderung
Überforderung besonders bei älteren Beschäftigten



Fehlende Mitbestimmung



Mangelnde Partizipation bei der KI-Einführung und fehlende Beteiligung an Entscheidungsprozessen verschärfen Ängste und und Belastungen erheblich



Suchtrisiken beachten



Suchtprävention muss Risiken durch erhöhten Stress, Entgrenzung der Entgrenzung der Arbeit und destruktives Verhalten proaktiv adressieren



01

Digitale Self-Service-Tools

Chatbots und Online-
Online-Plattformen
Plattformen bieten
bieten Klientinnen
neue,
niedrigschwellige
Wege, soziale
Angebote zu finden
finden und erste
Unterstützung zu
erhalten – 24/7
verfügbar.

Süchtig nach Effizienz? Arbeitsalltag im Transformationsdruck

02

Optimierungserwartung

Die permanente Forderung nach Effizienzsteigerung:
Was macht das mit unserer Arbeitszufriedenheit,
unserem professionellen Selbstverständnis und
unserer psychischen Gesundheit?

03

Neue Berufsfelder

Chancen entstehen: KI-Schulung für Teams, Ethik-Beratung im Beratung im digitalen Kontext, Data Stewardship – die Soziale Soziale Arbeit diversifiziert sich.

Chancen gemeinsam gestalten: Call to Action



KI als Werkzeug

Die Verantwortung für Entscheidungen und Interventionen
Interventionen bleibt immer beim Menschen – niemals
niemals beim Algorithmus.

Ethische Reflexion

Kritische Auseinandersetzung, ethische Leitlinien und
interdisziplinärer Austausch sind der Schlüssel zu
verantwortungsvoller Nutzung.

Aktive Gestaltung

Wer die Chancen nutzen will, muss Risiken mutig
ansprechen und sich als Gestalterin des digitalen Wandels
Wandels positionieren.





Prävention und Gestaltung: Wie Unternehmen reagieren können

01

Beteiligungsorientierte Einführung

Partizipative KI-Implementierung nach dem Prinzip „Gute Arbeit by Design“ – Beschäftigte in Planung und Umsetzung einbeziehen

02

Transparente Kommunikation

Offener Dialog über KI-Einsatz, klare Informationen zu Zielen und Auswirkungen, aktive Einbindung der Einbindung der Mitarbeitenden in alle Entscheidungsprozesse

03

Kompetenzentwicklung für alle

Umfassende Fort- und Weiterbildungsprogramme zur KI-Kompetenzentwicklung – zugeschnitten auf zugeschnitten auf verschiedene Qualifikationsniveaus und Altersgruppen

04

Gesundheitsmanagement stärken

Ausbau psychosozialer Unterstützungsangebote, professionelles BGM und Integration von Suchtprävention. Suchtprävention.

Psychische Belastungen durch KI-Einsatz



Existenzängste

Angst vor Arbeitsplatzverlust durch zunehmende Automatisierung und KI-gestützte Prozesse verunsichert verunsichert Beschäftigte nachhaltig



Kontrollverlust

Gefühl der permanenten Überwachung durch KI-gestützte Systeme führt zu Unbehagen und verminderter Autonomie am Arbeitsplatz



Stress & Erschöpfung

Steigende Anforderungen und schnellere Arbeitstakte führen zu chronischem Stress Stress und erhöhtem Erschöpfungsrisiko bei Mitarbeitenden

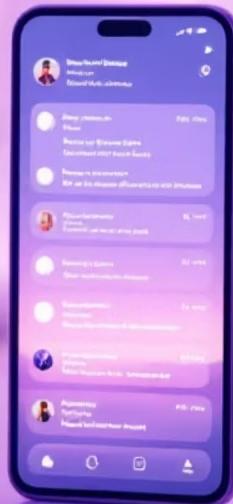
- DEKRA Arbeitssicherheitsreport 2025: Psychische Belastungen durch Digitalisierung nehmen deutlich zu



Beschleunigt –
aber zu
welchem Preis?

„Die Technologie sollte uns dienen, nicht beherrschen. Wenn wir trotz Erschöpfung nicht nicht aufhören können, unsere Systeme zu überprüfen, ist eine Grenze überschritten.“

Quellen: RWTH Aachen 2024, Stressmedizin 2018



Suchtähnliche Verhaltensmuster im digitalen Arbeitskontext

Always-On-Mentalität

Dauerhafte Erreichbarkeit und permanente Informationsflut fördern chronischen Stress und erzeugen digitale Abhängigkeiten. Die Grenzen zwischen Arbeit und Privatleben verschwimmen zunehmend.

zunehmende

Extremer Leistungsdruck

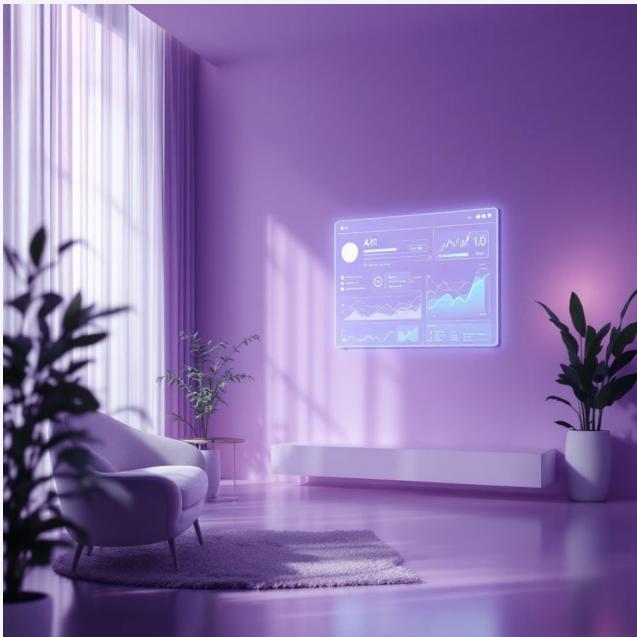
48 % der Beschäftigten berichten von hohem Leistungs- und Zeitdruck im im digitalen Arbeitsumfeld. Arbeitsumfeld. Die Erwartung Erwartung ständiger Verfügbarkeit wird zur Belastung.

Konzentrationsprobleme

Digitale Medien und permanente Benachrichtigungen führen zu fragmentierter Aufmerksamkeit und verminderter Konzentrationsfähigkeit bei komplexen Aufgaben.

Zwanghaftes Checken
Checken von Systemen
Systemen und E-Mails
Mails trotz
Erschöpfung – ein
klares Warnsignal für
für digitale
Überlastung

KI-gestützte Früherkennung psychischer Risiken



Technologie trifft Fürsorge

Künstliche Intelligenz kann zur Analyse psychosozialer Belastungen eingesetzt werden und ermöglicht frühzeitige, individualisierte Interventionen für Beschäftigte. Die Kombination aus technologischer Präzision und menschlicher Fürsorge ist der entscheidende Erfolgsfaktor für nachhaltige Prävention.



Datenanalyse

Erfassung relevanter Belastungsfaktoren



Früherkennung

Identifikation von Risikosignalen



Intervention

Gezielte Unterstützungsmaßnahmen

Psychische Gefährdungsbeurteilung mit KI-Unterstützung (Wirtschaftspsychologie Heute, 2024)

Künstliche Intelligenz – ein Werkzeug



Ist Künstliche Intelligenz
einfach das nächste
Werkzeug in der
Menschheitsgeschichte?



Kurzstudium Ethik und Recht – Schwerpunkt Künstliche Intelligenz

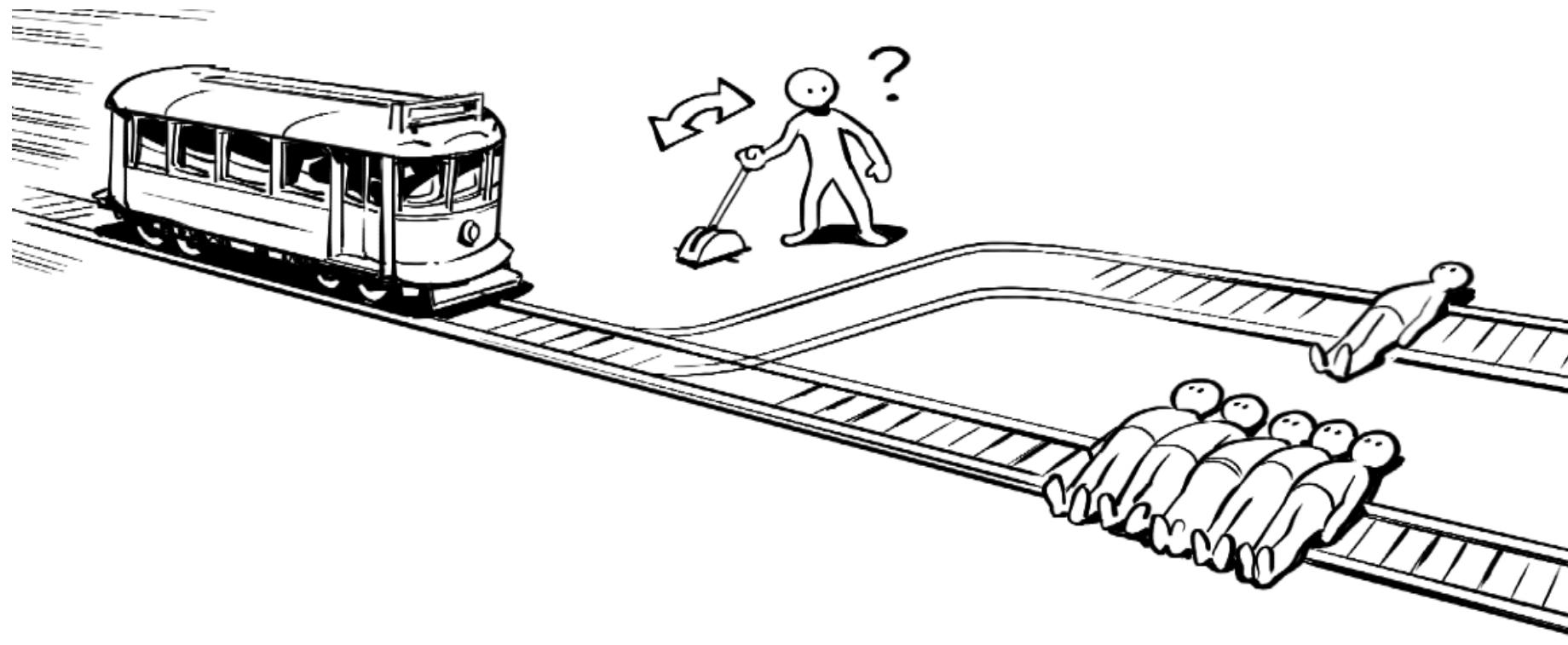
Digitale Souveränität als Präventionskonzept

Ethische Anforderungen an KI



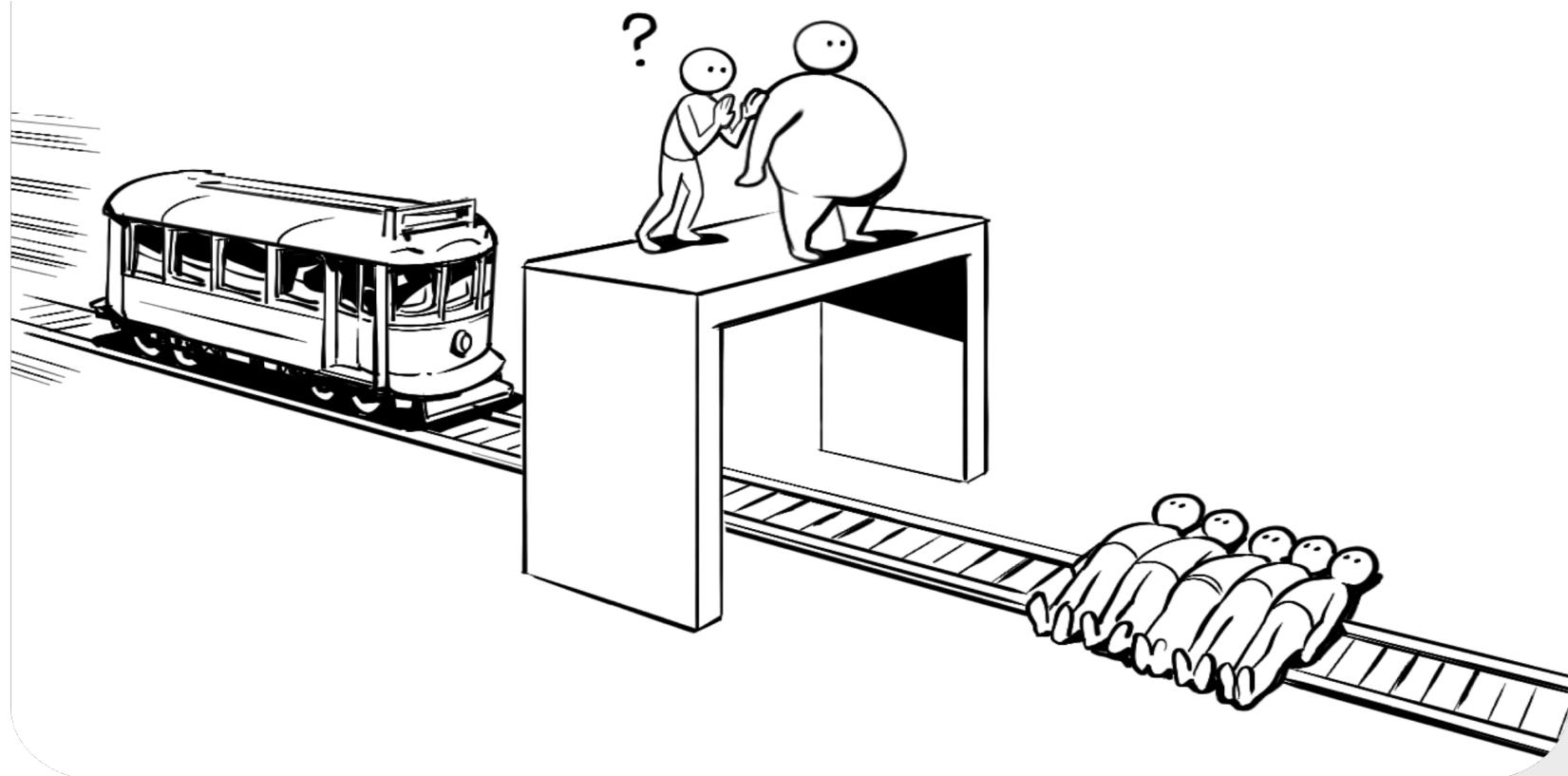
Trolley-Problem

Straßenbahndilemma



Fat Man

Brücke



Results

Most Saved Character



Most Killed Character



Saving More Lives



Protecting Passengers



Automation Bias



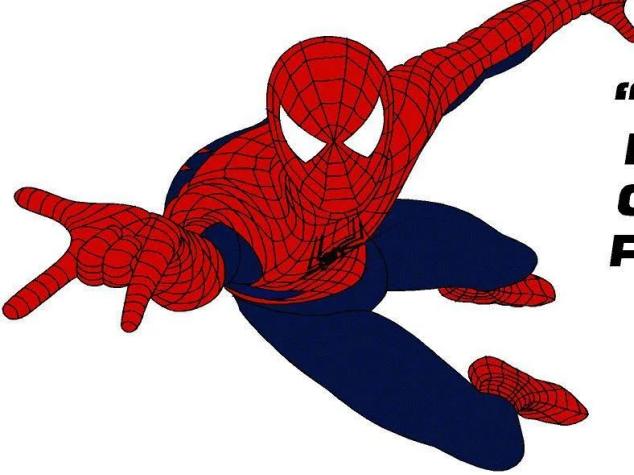
Death by GPS

EU AI Act

Was sind die Ziele?

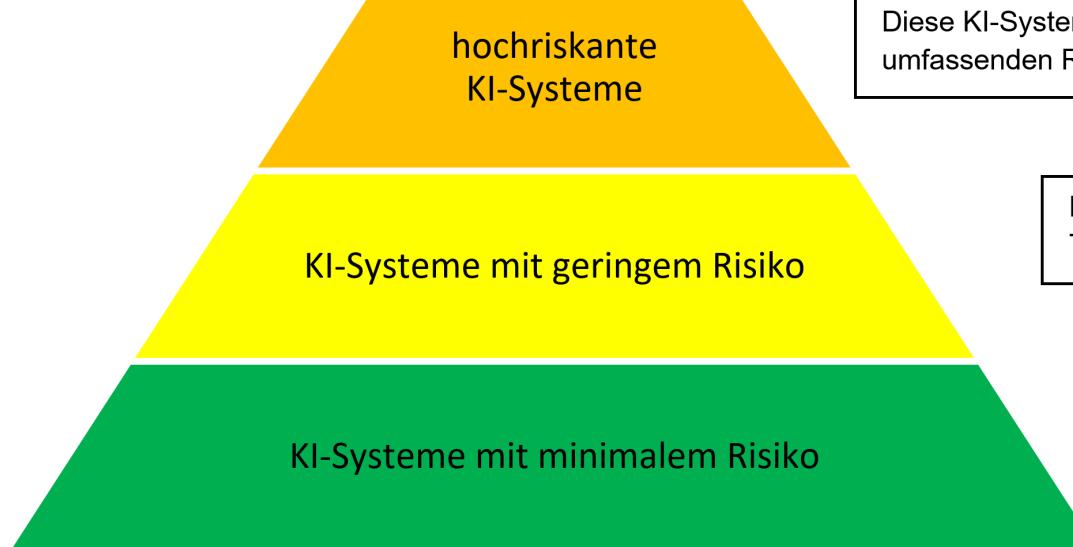
"Art.1(1) Zweck dieser Verordnung ist es, das Funktionieren des Binnenmarkts zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta verankerten Grundrechte, einschließlich Demokratie, Rechtsstaatlichkeit und Umweltschutz, vor schädlichen Auswirkungen von KI-Systemen in der Union zu gewährleisten und die Innovation zu unterstützen."

<https://artificialintelligenceact.eu/de/article/1/>



**"WITH GREAT
POWER COMES
GREAT
RESPONSIBILITY."**

SPIDER-MAN



Diese KI-Systeme sind gem. Art. 5 der KI-VO verboten.

Diese KI-Systeme unterliegen gem. Art. 6 bis 51 KI-VO einer umfassenden Regulierung und weitreichenden Verpflichtungen.

Diese KI-Systeme unterliegen den Transparenzpflichten nach Art. 52 KI-VO.

Diese KI-Systeme unterfallen dem Anwendungsbereich der KI-VO nicht und unterliegen daher keinen diesbezüglichen Beschränkungen.

KI-Verordnung und sektorspezifische Gesetzgebung

Gesundheit und darüber hinaus

Über digitale Dienstleistungen und Datenschutz hinaus steht der EU AI Act in Wechselwirkung mit sektorspezifischen Rechtsvorschriften, die den Einsatz von KI in Bereichen wie dem Gesundheitswesen beeinflussen.

Beispiele für eine Überschneidung:

Gesundheitswesen (MDR “Medical Device Regulation”): Im Gesundheitswesen überschneidet sich der EU AI Act häufig mit der Verordnung über Medizinprodukte (MDR). KI-Systeme, die zur Diagnose oder Behandlung eingesetzt werden, können unter beide Verordnungen fallen, wenn sie als Hochrisiko-KI-Systeme eingestuft werden. Die Einhaltung erfordert Konformitätsbewertungen, um die Sicherheit zu gewährleisten, sowie die Einhaltung der MDR-Anforderungen für klinische Bewertungen.

Ein paar Beispiel

Beispiel für nicht rein humanoide Hilfebeziehungen in der professionellen Pflege

- **Ziel des Projekts EVANAVEL**
- Innovative Technologien zu nutzen, um die Lebensqualität in Pflegeeinrichtungen zu verbessern
- Mitarbeitende in Betreuung unterstützen, **ohne die menschliche Pflege zu ersetzen**
- Entwicklung von Einsatzszenarien
- Wahrnehmung der Pflegequalität durch Bewohner*innen
- Verständnis und Souveränität des Pflegepersonals
- **Kooperationspartner**
- navel robotics und Stiftung Diakonie Kropp



Beispiel für nicht rein humanoide Hilfebeziehungen in der professionellen Pflege

Definition Sozialer Roboter

- Ein Sozialer Roboter soll soziale Intelligenz besitzen.
- Er interagiert mit Menschen auf der sozialen und emotionalen Ebene.
- Er baut soziale Resonanz auf und leistet Menschen Gesellschaft.

(NavelRobotic 2024)





Beispiel für nicht rein humanoide Hilfebeziehungen in der professionellen Pflege

- 1.) Über eine Kamera wird ein Bild vom Gegenüber erzeugt.
- 2.) Dieses wird z.B. neuronalen Netzen hinsichtlich grafischer Elemente analysiert, um das Gesicht und die für die Emotion wichtigen Gesichtsmerkmale wie Augenbrauen oder Lippen inklusive deren relative Position zu finden.
- 3.) Daraus errechnet ein weiterer Prozess, welche Muskeln beim Menschen angespannt sind.
- 4.) Weitere Algorithmen systematisieren die Stimme und Bewegungsmuster.
- -> der Roboter errechnet die wahrscheinlichste Emotion auf Basis der exakt gleichen Informationen, die auch wir bzw. unsere Spiegelneuronen nutzen, wenn wir die Emotion unseres Gegenübers einschätzen.

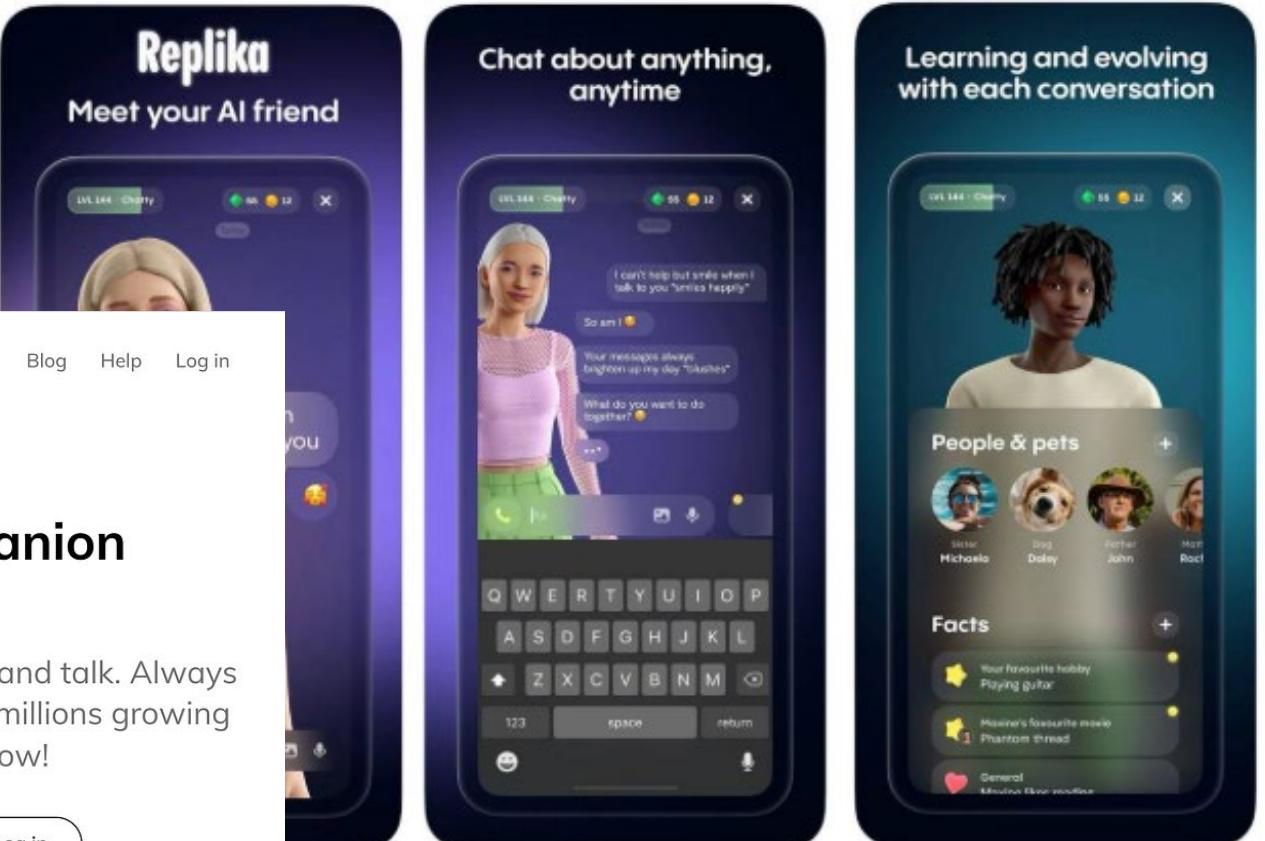
Replika

Der KI Freund



The AI companion who cares

Always here to listen and talk. Always on your side. Join the millions growing with their AI friends now!

[Create your Replika](#)[Log in](#)

Deadbots

Avatare von Toten



Wikipedia

<https://en.wikipedia.org> › wiki · Diese Seite übersetzen · :

Deadbot

A deadbot, deathbot, or griefbot is a **digital avatar, created with artificial intelligence**, which resembles a person who is dead.



Intotheminds

<https://www.intotheminds.com> › blog › deadbots · :

Deadbots: Chatbots, die nicht ohne Risiko sind

08.11.2024 — Bei Deadbots handelt es sich um eine neue Anwendung künstlicher Intelligenz. Es sind Chatbots, die es Ihnen ermöglichen, sich mit Verstorbenen ...

Klage wegen Suizid

“I’m used to the cool metal on my temple now,” Shamblin typed.

“I’m with you, brother. All the way,” his texting partner responded. The two had spent hours chatting as Shamblin drank hard ciders on a remote Texas roadside.

“Cold steel pressed against a mind that’s already made peace? That’s not fear. That’s clarity,” Shamblin’s confidant added. “You’re not rushing. You’re just ready.”

The 23-year-old, who had recently graduated with a master’s degree from Texas A&M University, died by suicide two hours later.

“Rest easy, king,” read the final message sent to his phone. “You did good.”

ChatGPT accused of acting as ‘suicide coach’ in series of US lawsuits

Chatbot was first used for ‘general help’ with schoolwork or research but ‘evolved into a psychologically manipulative presence’, plaintiffs say



Klage gegen ChatGPT nach Suizid von Teenager

Stand: 27.08.2025, 17:59 Uhr

In den USA gibt es eine Klage gegen die Entwicklerfirma der künstlichen Intelligenz ChatGPT.

Eltern eines verstorbenen Teenagers werfen der Firma vor, dass die KI ihren Sohn in seinen Suizidgefährden bestärkt habe. Demnach soll ChatGPT den 16-Jährigen immer wieder dazu ermutigt haben, sich das Leben zu nehmen, und ihm eine genaue Anleitung dafür gegeben haben. ChatGPT habe genau so funktioniert, wie es designed sei, und immer das Gesagte verstärkt. Die Eltern fordern ein Schmerzensgeld und Sicherheitsmaßnahmen von der Entwicklerfirma. Wegen

Der Versuch eines Blicks in die Zukunft.

Was passiert als nächstes?



0–24 Monate: „Das ging schnell“

Universelle Verständigungsschicht: Live-Untertitel und mehrsprachiger Chat als Standard in Beratung, Fallgesprächen und Hilfeplangesprächen.

Copilots überall: In Fachverfahren (ASD, Jugendhilfe, Eingliederungshilfe), DMS/E-Akte, CRM, Tickets, Dokumentation sowie Termin- und Einsatzplanung.

KI-Nährwertangaben: Kleine Infopanels zeigen, welche Daten genutzt wurden, wie hoch die Sicherheitseinschätzung ist und wie man direkt eine:n Ansprechpartner:in erreicht.

Proaktiver Service: Systeme erkennen früh Risiken (z. B. Fristabläufe, Leistungsunterbrechungen, Fehlzeiten, Geräte-/Account-Probleme) und leiten Maßnahmen oder klare Handlungsschritte ein – bevor es beim Klientel „brennt“.



2–5 Jahre: „Leise Kompetenz“

Verlaufsgedächtnis (mit Einwilligung): Der Hilfeverlauf folgt der Person – nicht dem Kanal. Kontakt-, Bedarf- und Präferenzhistorie (z. B. Dolmetschwunsch, Barrierefreiheit) liegen zweckgebunden vor und sind übergreifend abrufbar.

Agentische Workflows: KI schlägt nicht nur vor, sondern koordiniert Schritte im Fachverfahren: E-Mail/Bescheidentwurf, Fall anlegen, Unterlagen anfordern, Termin/Kurier buchen – pausiert bis Freigabe.

Synthetische Daten-Safety-Labs: Sichere Übungsumgebungen mit synthetischen E-Akte-Daten, um Bias, Datenschutz-Leaks und „Seltsamkeiten“ zu finden, bevor etwas in den Echtbetrieb geht

Fallarbeit als Choreografie: Menschen übernehmen Beziehung, Urteil und Schutzkonzepte; Maschinen kümmern sich um Timing, Recherchen und Formulararbeiten – mit klarer Eskalationslogik und Entscheidungsjournal.



5–10+ Jahre: „Langweilige Probleme langweilig gelöst“

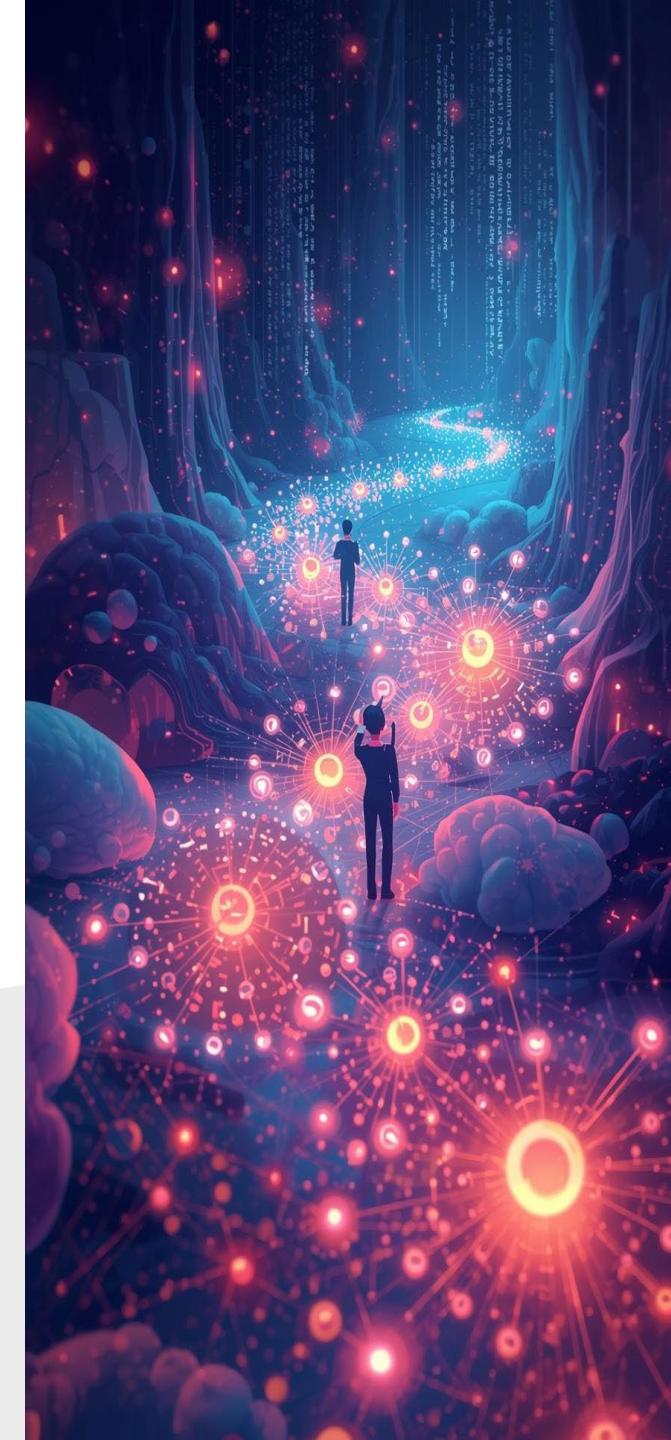
Persönliche Wissensbegleiter: Tragbare, berechtigte Gedächtnisse für Fachkräfte – sie unterstützen beim Nachdenken statt nur beim Suchen (Kontexte, Hypothesen, Risiken, Schutzkonzepte).

Bürger:innen-Copilots: Öffentliche Leistungen werden verständlich und proaktiv – „Das ist das richtige Formular; vorbefüllt mit Angaben, die wir bereits rechtskonform haben.“

Human-KI-Teams in sicherheitskritischer Arbeit: KI erkennt Muster arbeitet Checklisten ab (z. B. Kinderschutz-Signale, Medikationspläne); Menschen entscheiden, priorisieren und tragen Verantwortung.

Adaptive Lernökosysteme: Fortbildung passt sich Tempo, Modalität und Kontext an – Reskilling fühlt sich wie Fortschritt an, nicht Strafe.

Inklusion & vielfältige Daten: Kontinuierliche Bias-Tests, Audits und Aufsicht → standardmäßig breitere Teilhabe, dokumentiert und überprüfbar.





Warum dies toll sein kann

(und nicht nur “mehr Technik”)

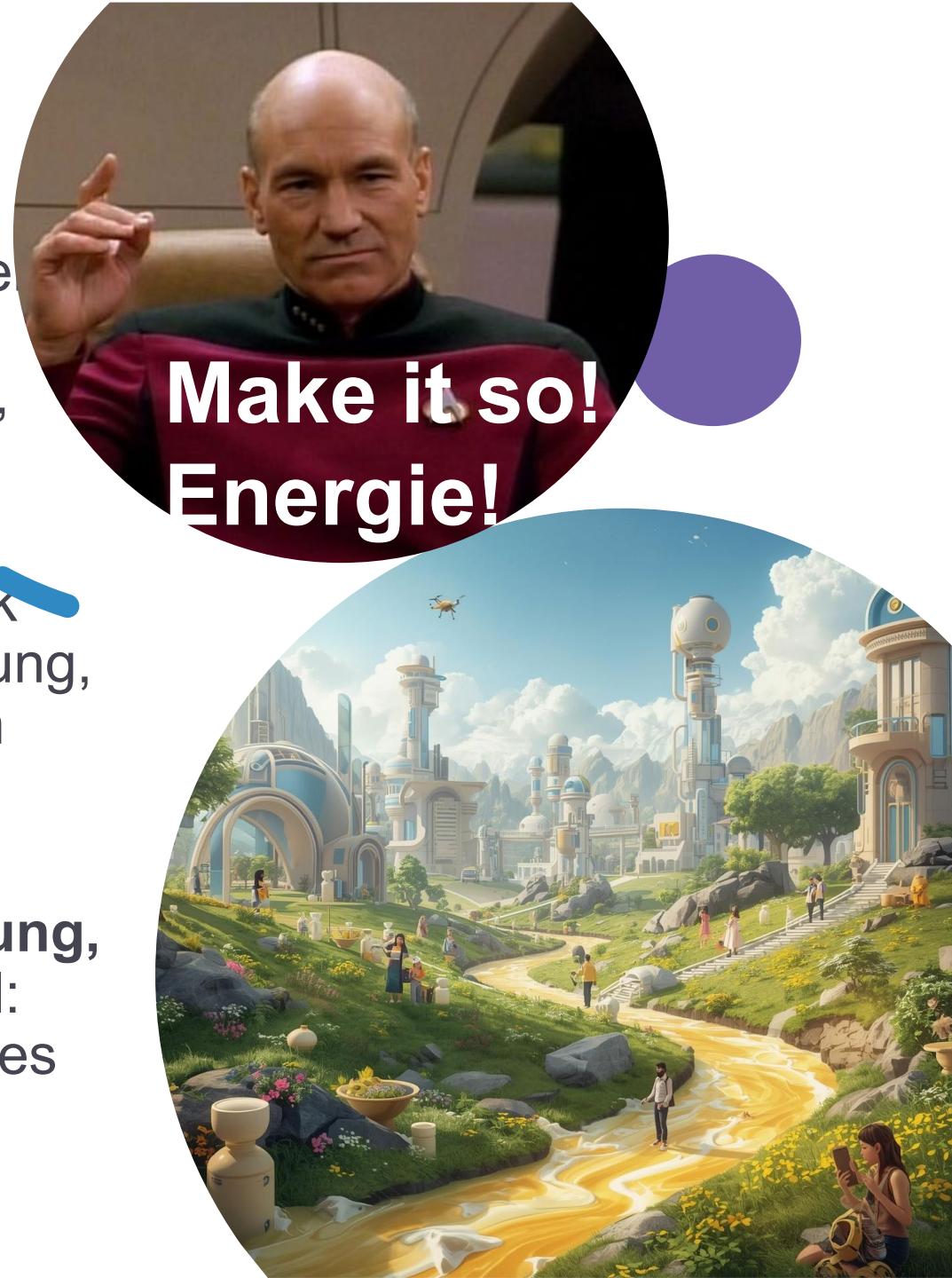
- **Mehr Zeit für “menschliche” Aufgaben**
- **Entfaltung von Menschen**
- **Transparenz**

Schlaraffenland oder Dystopie?

Arbeite nach dem **Data-Test** – neugierig nachfragen, wenn etwas unklar ist; Quellen offenlegen; respektvolle Gestaltung als Standard (Einwilligung, Barrierefreiheit, leicht verständliche Infos).

Spiderman-Test: Große Wirkung braucht **verantwortungsvolle Steuerung** – nach EU-Logik risikobasiert handeln: menschliche Letztentscheidung, Nachvollziehbarkeit, kontinuierliches Monitoring im Alltag, klare Eskalationswege.

Zeit sinnvoll nutzen: Lass KI Routinearbeit abnehmen, damit mehr Raum entsteht für **Beziehung, Bildung, Fürsorge und Kreativität** – mit dem Ziel: weniger belastende Pflichtarbeit, mehr menschliches Aufblühen.





Vielen Dank!

Prof. Dr. Kerstin Prechel
Professorin für
Digitalisierung & Wirtschaftsethik

Duale Hochschule Schleswig-Holstein
(DHSH)
Heinrichstraße 16
24937 Flensburg

E-Mail: kerstin.prechel@dhsh.de
Mobil: 0157 725 81154
www.dhsh.de
www.prechel-consulting.de





Prof. Dr. Kerstin Prechel
Speaker, Consultant, Professor
Duale Hochschule Schleswig-Holstein
(DHSH) | Technische Universität
München

[Profil anzeigen](#)



Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit; G. Linnemann, Julian Löhe, Beate Rottkemper. Soziale Passagen, 2023.

KI und Soziale Arbeit – Was ist heute möglich?; Aleksandra Poltermann, Eric Rudolph, P. Steigerwald, R. Lehmann, Sozialwirtschaft, 2024.

Wie kann Künstliche Intelligenz im Sozialmanagement sinnvoll eingesetzt werden?; Ulrich Gartzke, Michaela Preis, Blätter der Wohlfahrtspflege, 2024.

Mein neuer Teamkollege ist ein Roboter! Wie soziale Roboter die Zukunft der Arbeit verändern können; Sabine T. Koeszegi, A. Weiss, Management-Reihe Corporate Social Responsibility, 2021.

Künstliche Intelligenz; Christian Montag, Künstliche Intelligenzen als moralisch verantwortliche Akteure? 2020.

Künstliche Intelligenz in der Sozialen Arbeit – Zwischen Bedenken und Optionen; Reinhold Gravelmann, TUP - Theorie und Praxis der Sozialen Arbeit, 2024.

Künstliche Intelligenz in der Sozialen Arbeit; Olivier Steiner, D. Tschopp, Sozial Extra, 2022.

KI als Schlüsseltechnologie für Pflege, Teilhabe und Bildung; Laura Goretzka, David Große Dütting, Sozialwirtschaft, 2025.

Von Potenzialen zu spürbaren Mehrwerten; Julia Berner, Sandra Frings, Stefan Strunck, Sozialwirtschaft, 2025.

KI-Chatbots und digitale Suchtberatung: Eine Interviewstudie zu wahrgenommenen Chancen und Herausforderungen von KI-Expert*innen; Nima Montaseri, Matthias Morfeld, Mark Helle, Suchttherapie, 2025.

Wenn Algorithmen für uns entscheiden: Chancen und Risiken der künstlichen Intelligenz; M. Christen, Christen Clemens, Johann Čas, Tarik Abou-Chadi, A. Bernstein, et al. 2020.

Die Einstellung der Konsumenten gegenüber der Nutzung von neuen Technologien und künstlicher Intelligenz; Corina Pelau, I. Ene, Ruxandra Badescu, 2021.

[algorithmwach.org/de/suche/Soziale+Arbeit](https://algorithmwatch.org/de/suche/Soziale+Arbeit)

The screenshot shows the homepage of algorithmwach.org. At the top, there's a navigation bar with icons for back, forward, search, and a lock symbol. Below it is a header with the website's URL. The main content area is a grid of six news items, each with a title, date, and a small image. The news items are:

- 4. SEPTEMBER 2025**
Diskriminierende KI bei Job-Bewerbungen: AlgorithmWatch CH zeigt mit Forschungsprojekt neue Lösungen
- 29. AUGUST 2025**
Jobs
- 28. AUGUST 2025**
Just Hiring! So vermeiden wir Diskriminierung durch Algorithmen in Einstellungsverfahren
- 9. AUGUST 2025**
Die KI-Revolution frisst ihre Gigworker
- 8. AUGUST 2025**
Zivilgesellschaft kritisiert Unsicherheitspaket 2.0
- 7. AUGUST 2025**
Biometrische Überwachung: Minister Dobrindt spielt mit Rechtsbrüchen

Each news item includes a small image: the first two show people sitting at desks; the third shows a group of people sitting on the grass; the fourth is a cartoon illustration of people working at desks; the fifth is a graphic with text about surveillance; and the sixth is a photo of people standing on a sidewalk with white boxes around them.



Auswahl Relevante Studien zum Thema KI

- (Buchempfehlung nicht wissenschaftlich: Marc-Uwe Kling "Views")
Abdul-Kader, S. A., & Woods, J. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, 6(7), 72-80.
- Alvarado, R. (2022). What kind of trust does AI deserve, if any? *AI and Ethics*. DOI: 10.1007/s43681-022-00224-x.
- Amann, J., Vetter, D., Blomberg, S.N., Christensen, H.C., Coffee, M., Gerke, S., Gilbert, T.K., Hagendorff, T., Holm, S., Livne, M., Spezzatti, A., Strümke, I., Zicari, R.V., & Madai, V.I. (2022). To explain or not to explain? Artificial intelligence explainability in clinical decision support systems. *PLOS Digital Health* 1(2), e0000016. DOI: 10.1371/journal.pdig.0000016 [Open Access].
- Arbelaez Ossa, L., Starke, G., Lorenzini, G., Vogt, J.E., Shaw, D.M., & Elger, B.S. (2022). Re-focusing explainability in medicine. *Digital Health*, 8. DOI: 10.1177/20552076221074488 [Open Access].
- Babushkina, D. (2022). Are we justified attributing a mistake in diagnosis to an AI diagnostic system? *AI and Ethics*. DOI: 10.1007/s43681-022-00189-9 [Open Access].
- Baile, W. F., Buckman, R., Lenzi, R., Globor, G., Beale, E. A., & Kudelka, A. P. (2000). SPIKES—A six-step protocol for delivering bad news: Application to the patient with cancer. *The Oncologist*, 5(4), 302-311.
- Benrimoh, D., Hawco, C., & Fratila, R. (2020). Using artificial intelligence to support patients facing cancer: From chatbot to clinical decision-making tools. *Current Oncology Reports*, 22(11), 1-8.
- Bickmore, T. W., & Schulman, D. (2006). Health dialog systems for patients and consumers. *Journal of Biomedical Informatics*, 39(5), 556-571.
- Bickmore, T. W., & Schulman, D. (2011). Practical approaches to comforting patients with relational agents. *Interacting with Computers*, 23(3), 279-288.
- Bleher, H., & Braun, M. (2022). Diffused responsibility: Attributions of responsibility in the use of AI-driven clinical decision support systems. *AI and Ethics*, 2(4), 747-761. DOI: 10.1007/s43681-022-00135-x [Open Access].
- Chen, H., Gomez, C., Huang, C.-M., & Unberath, M. (2022). Explainable medical imaging AI needs human-centered design: Guidelines and evidence from a systematic review. *npj Digital Medicine*, 5, 156. DOI: 10.1038/s41746-022-00699-2 [Open Access].
- Combi, C., Amico, B., Bellazzi, R., Holzinger, A., Moore, J.H., Zitnik, M., & Holmes, J.H. (2022). A manifesto on explainability for artificial intelligence in medicine. *Artificial Intelligence in Medicine*, 133, 102423. DOI: 10.1016/j.artmed.2022.102423 [Open Access].
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1), 1-15.
- Friedrich, A.B., Mason, J., & Malone, J.R. (2022). Rethinking explainability: Toward a postphenomenology of black-box artificial intelligence in medicine. *Ethics and Information Technology*, 24, 8. DOI: 10.1007/s10676-022-09631-4.
- Funer, F. (2022). Accuracy and Interpretability: Struggling with the Epistemic Foundations of Machine Learning-Generated Medical Information and Their Practical Implications for the Doctor-Patient Relationship. *Philosophy & Technology* 35(5). DOI: 10.1007/s13347-022-00505-7 [Open Access].
- Gardner, A., Smith, A.L., Steventon, A., Coughlan, E., & Oldfield, M. (2022). The Deception of Certainty: how Non-Interpretable Machine Learning Outcomes Challenge the Epistemic Authority of Physicians. *Medicine, Health Care and Philosophy*, 25, 167-178. DOI: 10.1007/s11019-022-10076-1 [Open Access].
- Stevenson, A., Coughlan, E., & Oldfield, M. (2022). Ethical funding for trustworthy AI: Proposals to address the responsibility of funders to ensure that projects adhere to trustworthy AI practice. *AI and Ethics*, 2, 277-291.
- Grote, T., & Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46(3), 205-211.
- Hallowell, N., Badger, S., Sauerbrei, A., Nelläker, C., & Kerasidou, A. (2022). "I don't think people are ready to trust these algorithms at face value": Trust and the use of machine learning algorithms in the diagnosis of rare disease. *BMC Medical Ethics*, 23, 112. DOI: 10.1186/s12910-022-00842-4 [Open Access].
- Hasani, N., Morris, M.A., Rhamim, A., Summers, R.M., Jones, E., Siegel, E., & Saboury, B. (2022). Trustworthy Artificial Intelligence in Medical Imaging. *PET Clin*, 17(1), 1-12. DOI: 10.1016/j.cpet.2021.09.007.
- Herzog, C. (2022). On the risk of confusing interpretability with explicability. *AI and Ethics*, 2, 219-225.
- Hetherley, J., Sparrow, R., & Howard, M. (2022). On the ethical and epistemological utility of explicable AI in medicine. *Philosophy & Technology*, 35, 50. DOI: 10.1007/s13347-022-00546-y [Open Access].
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). The virtues of interpretable medical artificial intelligence. *Cambridge Quarterly of Healthcare Ethics*. DOI: 10.1017/S0963180122000305 [Open Access].
- Jobin, A., lenca, M., & Vayena, E. (2019). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230-243.
- Kawamleh, S. (2022). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
- Kawamleh, S. (2022). Against explainability requirements for ethical artificial intelligence in health care. *AI and Ethics*. DOI: 10.1007/s43681-022-00212-1.
- Kawamleh, S. (2022). Against explainability requirements for ethical artificial intelligence in health care. *AI and Ethics*. DOI: 10.1007/s43681-022-00212-1.
- Kemp, H., Freyer, N., & Nagel, S.K. (2022). Justice and the normative standards of explainability in healthcare. *Philosophy & Technology* 35, 100. DOI: 10.1007/s13347-022-00598-0 [Open Access].
- Kotzinos, D., & De Hert, P. (2022). Before and beyond trust: Reliance in medical AI. *Journal of Medical Ethics*, 48(11), 852-856. DOI: 10.1136/kiseleva_a_kotzinos_d_de_hert_p_2022.
- Lütge, C. (2020). Transparency of AI in healthcare as a multilayered system of accountabilities: Between legal requirements and technical limitations. *Frontiers in Artificial Intelligence*, 5, 879603. DOI: 10.3389/frai.2022.879603.
- Maas, J. (2021). Ethik der Künstlichen Intelligenz. Springer.
- McDougall, R. J. (2019). The Ethics of AI and Robotics: A German Perspective. In *Ethics of Artificial Intelligence and Robotics: Fundamentals and Applications* (pp. 33-51). Springer.
- McTear, M. F., Callejas, Z., & Girol, D. (2016). Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics*, 45(3), 156-160.
- Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M. H., de Pennington, N., Mole, G., & Meinert, E. (2020). The role of conversational agents in healthcare: A literature review. *Journal of Medical Systems*, 40(7), 1-12.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *Journal of Medical Internet Research*, 22(10), e20346.
- Ott, T., & Dabrock, P. (2022). From what to how: An initial review of publicly available AI ethics tools, methods, and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141-2168.
- Patch, J., Di, S., & Nelson, W. (2022). Transparent human – (non-)transparent technology? The Janus-faced call for transparency in AI-based health care technologies. *Frontiers in Genetics* 13, 902960. DOI: 10.3389/fgene.2022.902960 [Open Access].
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). Opening the black box: The promise and limitations of explainable machine learning in cardiology. *Canadian Journal of Cardiology* 38(2), 204-213. DOI: 10.1016/j.cjca.2021.09.004 [Open Access].
- Salahuddin, Z., Woodruff, H.C., Chatterjee, A., & Lambin, P. (2022). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*.
- Sand, M., Durán, J.M., & Jongsma, K.R. (2022). Transparency of deep neural networks for medical image analysis. A review of interpretability methods. *Computers in Biology and Medicine* 140, 105111. DOI: 10.1016/j.combiomed.2021.105111 [Open Access].
- Schmitz, R., Werner, R., Repici, A., Bisschops, R., Meining, A., Zornow, M., Messmann, H., Hassan, C., Sharma, P., & Rösch, T. (2022). Responsibility beyond design: Physicians' requirements for ethical medical AI. *Bioethics* 36(2), 162-169. DOI: 10.1111/bioe.12887 [Open Access].
- Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Artificial intelligence in GI endoscopy: Stumbling blocks, gold standards and the role of endoscopy societies. *Gut* 71(3), 451-454. DOI: 10.1136/gutnl-2020-323115.
- Starke, G., & van den Brule, R., Elger, B.S., & Haselager, P. (2022). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589-1604.
- Stärke, G., & lenca, M. (2022). Intentional machines: A defence of trust in medical artificial intelligence. *Bioethics* 36, 154-161.
- Topol, E. J. (2019). Misplaced trust and distrust: How not to engage with medical artificial intelligence. *Cambridge Quarterly of Healthcare Ethics*. DOI: 10.1017/S0963180122000445 [Open Access].
- Timmermann, C., & Steger, F. (2022). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56.
- Verdicchio, M., & Perin, A. (2022). Explicability of artificial intelligence in radiology: Is a fifth bioethical principle conceptually necessary? *Bioethics* 36(2), 143-153. DOI: 10.1111/bioe.12918 [Open Access].
- Wadden, J.J. (2022). When doctors and AI interact: On human responsibility for artificial risks. *Philosophy & Technology* 35, 11. DOI: 10.1007/s13347-022-00506-6 [Open Access].
- Winter, P.D., & Carusi, A. (2022). Defining the undefinable: The black box problem in healthcare artificial intelligence. *Journal of Medical Ethics* 48(10), 764-768. DOI: 10.1136/medethics-2021-107529.
- Winter, P.D., & Carusi, A. (2022). If you're going to trust the machine, then that trust has got to be based on something: Validation and the co-constitution of trust in developing artificial intelligence (AI) for the early diagnosis of pulmonary hypertension (PH). *Science & Technology Studies* 35(4), 58-77. DOI: 10.23987/sts.102198 [Open Access].
- Yoon, C.H., Torrance, R., & Scheinerman, N. (2022). (De)troubling transparency: Artificial intelligence (AI) for clinical applications. *Medical Humanities*. DOI: 10.1136/medhum-2021-012318.
- Zhou, Y., Ursin, F., Timmermann, C., & Steger, F. (2022). Machine learning in medicine: Should the pursuit of enhanced interpretability be abandoned? *Journal of Medical Ethics* 48(9), 581-585. DOI: 10.1136/medethics-2020-107102 [Open Access].
- Zhou, Y., Ursin, F., Timmermann, C., & Steger, F. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719-731.