



Algorithmen der Vielfalt

KI als Wegbereiterin für echte Inklusion?

FH-Prof.ⁱⁿ Mag.^a Dr.ⁱⁿ Martina Gaisch

<fnma>

FORUM NEUE MEDIEN
IN DER LEHRE AUSTRIA

HAGENBERG | LINZ | STEYR | WELS



UNIVERSITY
OF APPLIED SCIENCES
UPPER AUSTRIA

Fluch oder Segen?



Dual Use – Collingridge-Dilemma



Gaisch & Mader (2025). AI Ethics & Human Factors in AI. Next Generation Digital. Excellence Edition. Wien

Künstliche Intelligenz

sichtet



KI als Sparring-Partnerin
als personalisierte Assistenz

BRAUCHT DRINGEND

- Human Factor /Human-in-the-Loop
- Faktencheck
- Kritische Bewertung
- Überprüfung der Quellenqualität
- Wissen über Halluzination
- Sensibilität über Bias



A fool with a tool is still a fool

Werkzeug allein macht keinen Unterschied – wenn nicht kompetent & informiert eingesetzt

- Kontextwissen & Wissen über Grenzen von KI
 - Verantwortung & moralische Instanz
 - Urteilsvermögen & menschliche Kontrolle

Gaisch & Mader (2025). AI Ethics & Human Factors in AI. Next Generation Digital. Excellence Edition. Wien

Künstliche Intelligenz

dichtet



Für KI unerreichbar:

- wahre **Originalität & Improvisation** aufgrund von menschlichen **Emotionen**, **Kontextverständnis**, persönlicher Erfahrungen
- **Emotionale Tiefe & Intuition**, die menschliche Kreativität einzigartig machen.
- **Kulturelles & soziales** Verständnis
- Ethische **Urteilsfähigkeit**
- Verständnis über moralische Implikation von KI-Kreationen

Künstliche Intelligenz



richtet

KI als Roborichterin

GEHT NICHT OHNE



soziale Kompetenz (§ 9 Nr. 4 DRiG)
Kontextwissen – Handeln nach
„bestem Wissen und Gewissen“ (§ 38 Abs 1 DRiG).

in Europa:

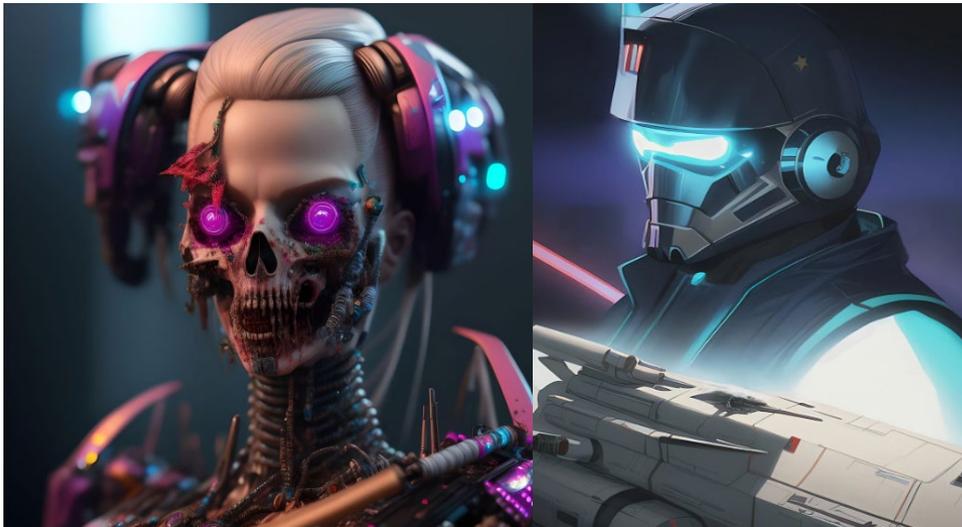
Rechtsprechung verboten
von **natürlicher** Person auszuüben
OLGA von IBM für Routineaufgaben

in USA:

Risikobewertung **COMPAS**
Rückfallrisiko von Straftätern
Afro-Amerikanische Personen wurden
systematisch diskriminiert

Künstliche Intelligenz

vernichtet



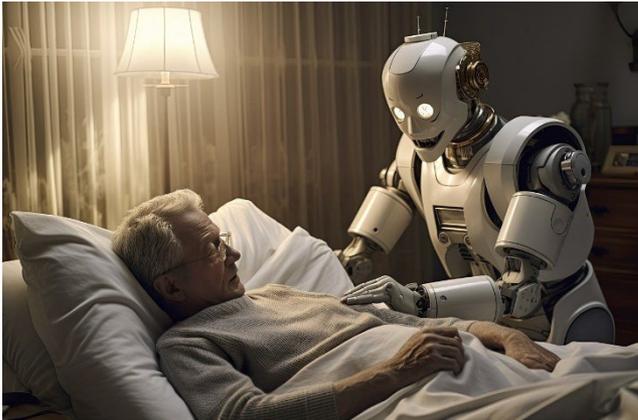
KI ohne Regulativ
(Gesetze, Ethik-Richtlinien, Compliance,
AI-Act)

FÜHRT ZU

Cyberkriminalität
Hate Speech
Deepfake
Manipulation – Social Bots
Spaltung der Gesellschaft

Welche KI-Systeme gehören auf den Prüfstand

die über Menschen
entscheiden



die über Ressourcen,
die Menschen
betreffen, entscheiden



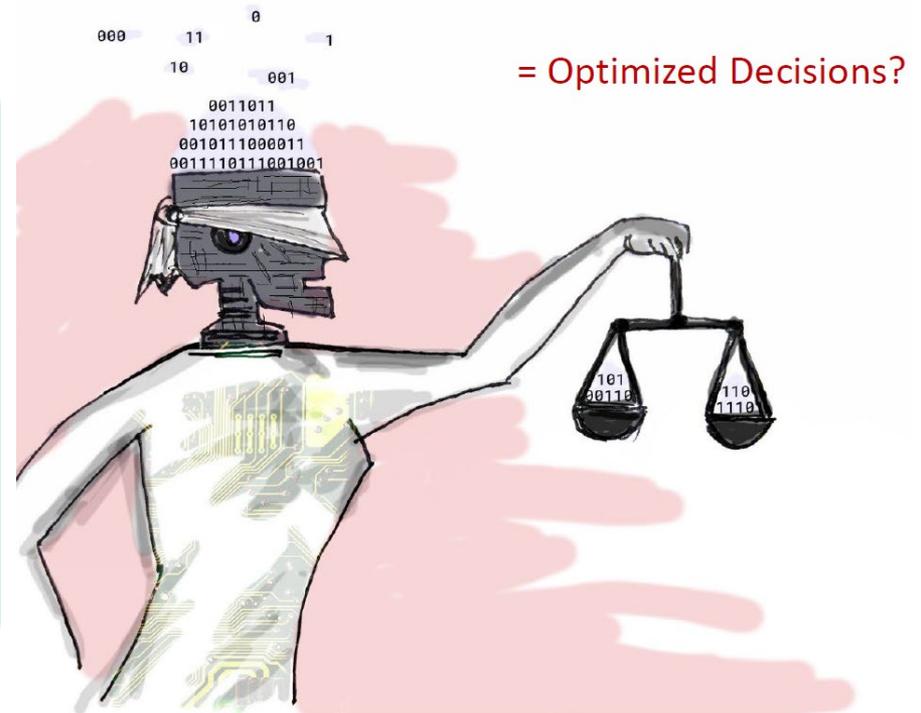
die über
gesellschaftliche
Teilhabe von Personen
entscheiden



Verwendung von Algorithmen für „bessere Entscheidungen“

- um ethische Herkunft **vorherzusagen**
- um Terroristen zu **detektieren**
- um Bewerbungen zu **bewerten**
- um über Entlassungen zu **entscheiden**
- um Urteile zu **fällen**
- um Diagnosen zu **unterstützen**
- um Gesundheitsleistungen zu **steuern**

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4), 835–850.



Macht KI die Welt inklusiver?



Automation Bias

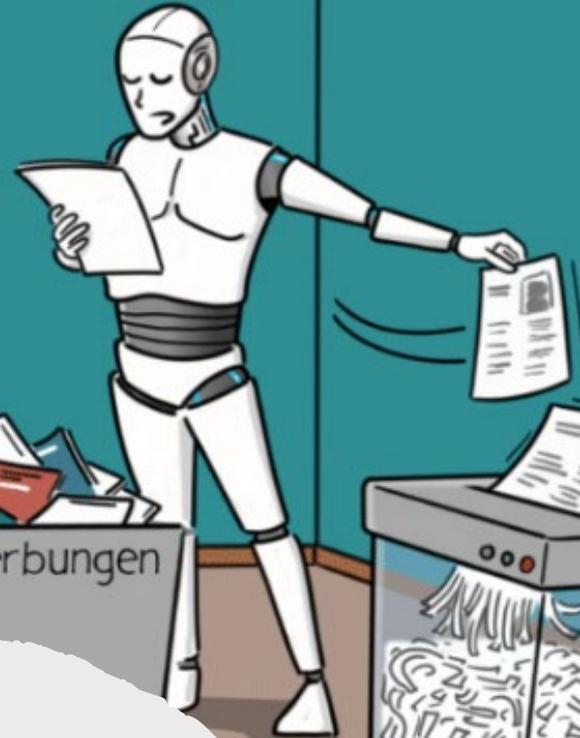


- **Blindes** Vertrauen in Technologie
- **Unkritischer** Akzeptanz von KI-Ergebnissen
- Automatisierung & Effizienz führt nicht automatisch zu Inklusion
- Inklusion braucht eine **bewusste** Gestaltung & Berücksichtigung **diverser** Bedürfnisse
- Unterschätzung kultureller/sozialer Vielfalt
- **Irrglaube**: KI sei **universelles** Werkzeug und funktioniert für alle gleich gut

WARUM SIE KEINEN KREDIT ERHALTEN?
WEIL MEIN ENTWICKLER DAHINTEN IN EINER
FILTERBLASE STECKT UND ICH DESWEGEN
MIT EINER VERZERRTEN DATENBASIS
ARBEITEN MUSSTE.



Zu alt,
zu dunkelhäutig,
zu weiblich...



objektiv?
neutral?
universell?

Spiegel unserer Gesellschaft



Garbage In – Garbage Out (GIGO)



Qualität der **Ergebnisse** hängt von der
Qualität der **Eingabedaten** ab

Datenlage:

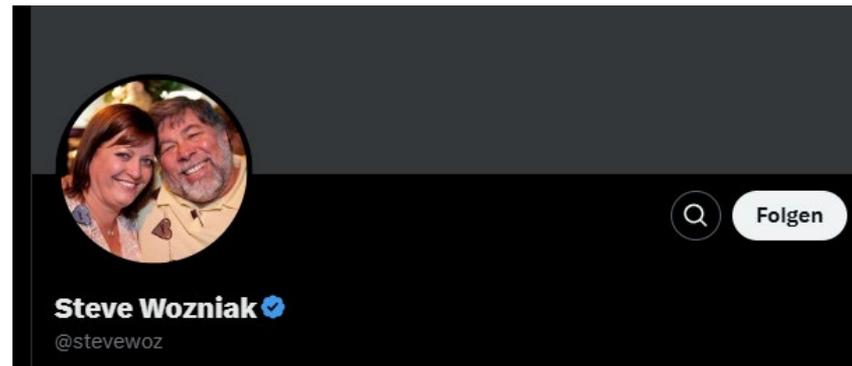
- Training mit **schlechten** Daten
- fehlerhaft, unvollständig, verzerrt
- **Bias** in den Daten
- Fehlende **Repräsentativität**
- Fehlende **Anpassung** an lokale
Kontexte

KI-System – problematische Ergebnisse

1. Kreditvergabe Apple-Mitbegründer **Wozniak**
2. Chatbot **Tay** 2016
3. **Amazon** Recruitment Tool 2014
4. Gesichtserkennung
5. Predictive Policing (**COMPAS**)



Fall 1: Kreditvergabe



Ein bemerkenswertes Beispiel ist der Fall von Janet Hill, der Ehefrau des Apple-Mitbegründers Steve Wozniak. Trotz identischer finanzieller Voraussetzungen und eines gemeinsamen Bankkontos erhielt Hill ein deutlich niedrigeres Kreditlimit als Wozniak – ein Beispiel dafür, wie algorithmische Entscheidungen bestehende Geschlechterungleichheiten nicht nur widerspiegeln, sondern verschärfen können (Wozniak 2019).

Fall 2: Chatbot Tay (Microsoft)



- der experimentelle Bot lernte durch menschliche Interaktion (März 2016)
- Tay wurde **in nur 24 Stunden** zum Hitlerbot
- musste nach einem Tag vom Netz genommen werden
- leugnete Holocaust und verbreitete Hasstiraden, Verschwörungstheorien
- agierte als Rassist, Sexist, Homophob & verbreitete extremistischen Content

Vom Hipster-Mädchen zum Hitler-Bot

Nicht einmal einen Tag twitterte der neue Chatbot von Microsoft. Dann lehrten ihn Twitternutzer Rassismus, Sexismus und Hass.

Von **Teresa Slickert**
24.03.2016, 17:13 Uhr



"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI



Microsofts KI findet Hitler besser als Obama

Veröffentlicht am 24.03.2016 | Lesedauer: 3 Minuten

Peinliche Panne für Microsoft künstliche Intelligenz „Tay“ auf Twitter: Der Chat-Bot lernte vor allem Rassismus und Pöbeleien vom Internet, was im Aufruf zum Völkermord gipfelte.

Chatbot "Tay" blamierte Microsoft mit rassistischen Entgleisungen

Experiment auf Twitter wurde zum Debakel

25. März 2016, 18:39

Microsoft hat eine Chatbot-Software von Netz nehmen müssen, nachdem Nutzer sie durch Tricks massenhaft zu rassistischen Entgleisungen verleiten konnten. Das am Mittwoch eingeschaltete Programm mit dem Namen "Tay" sollte mit Hilfe einfacher Antwort-Algorithmen über Twitter den Anschein einer Konversation mit Menschen aufrechterhalten, aber auch aus Unterhaltungen lernen.

Anpassungen



Microsoft

Twitter-Nutzer machen Chatbot zur Rassistin

Tay, ein Chatbot von Microsoft mit künstlicher Intelligenz, sollte im Netz lernen, wie junge Menschen reden. Nach wenigen Stunden musste der Versuch abgebrochen werden.

Von **Patrick Beuth**
24. März 2016, 15:13 Uhr

Microsoft-Programm "Tay"

Rassistischer Chat-Roboter: Mit falschen Werten bombardiert

3. April 2016, 15:25 Uhr | Lesezeit: 3 min



Erst plapperte "Tay" über Welpen. Dann verleugnete der Chatbot den Holocaust. (Foto: Screenshot: Twitter/Tayandyou)

Microsoft hat den lernfähigen Chat-Roboter "Tay" auf Twitter losgelassen. Bis Nutzer das Programm rassistische Propaganda lehrten.

Holocaust verleugnet: Microsoft entschuldigt sich für Chatbot "Tay"

Experimenteller Twitter-Bot fiel mit rassistischen und sexistischen Nachrichten auf

26. März 2016, 11:08

Nur 16 Stunden war Microsofts Twitter-Bot "Tay" aktiv. Nachdem die Chatbot-Software unter anderem Hitler angepriesen hatte, den Holocaust verleugnete, Schwarze sowie Mexikaner beleidigte und Feminismus mit Krebs verglich, war es dem Konzern genug. In einem Blog-Eintrag gab Microsoft bekannt, dass man diese Aussagen "zutiefst bedauere" und sich dafür entschuldige. Ein Großteil der Tweets ist mittlerweile gelöscht.



Foto: Screenshot/WebStandard

Fall 3: Amazon Recruitment Tool 2014

executed
captured



Fall 4: Gesichtserkennungssoftware

- Bias gegenüber **Hautfarben**
- Bias gegenüber **Frauen**

Fighting Bias in Algorithms

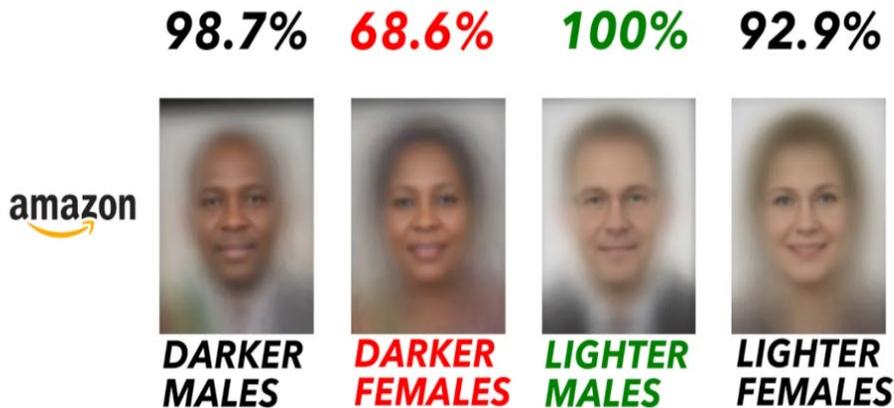
Joy Buolamwini

https://www.youtube.com/watch?v=UG_X_7g63rY



Seite | 21

Eine Frage der Datenlage



Amazon Rekognition Performance on Gender Classification

Algorithmic Justice League (AJL) (US)

Human Bias hat viele Gesichter
Intersektionalität

Rassismus

Skin Care, Seifenspender (Sensor reagiert nicht)

Sexismus

Lensa-App generiert sexualisierte Avatare in Anime-Style auch für Frauen, leicht bekleidet – Datenmaterial mit pornografischem Inhalt

Klassismus

Bei Beurteilung der Kreditwürdigkeit werden Minderheiten oder Menschen mit weniger Einkommen ungenauer berechnet, da es wenig Referenzdaten gibt.

Ableismus

Menschen mit körperlicher Behinderung werden im Straßenverkehr kaum erkannt, da ihre Bewegungsmuster nicht verstanden/nicht erhoben werden.

Fall 5: Predictive Policing



COMPAS Algorithmen

Verwendung vor der Verhandlung, bei Verurteilung, Bewährungsanhörung



Zweck

Prognostiziert die Wahrscheinlichkeit einer erneuten Straftat



Methode

Erzeugt Risikobewertungen auf der Grundlage von über 100 Faktoren aus einem detaillierten Fragebogen.



Kontroverse

Kritisiert wegen Rassenungleichheiten und mangelnder Transparenz

Risk Assessment Beispiel



Frühere Straftaten:

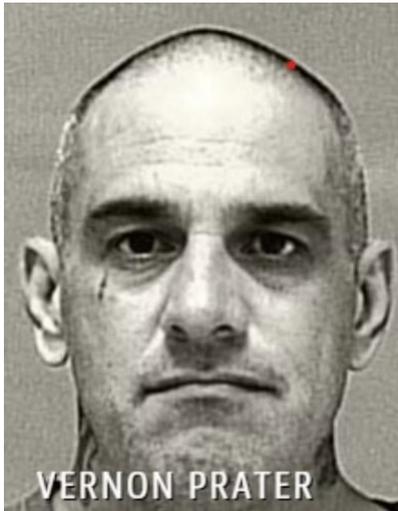
2 bewaffnete Raubüberfälle
1 versuchter bewaffneter Überfall



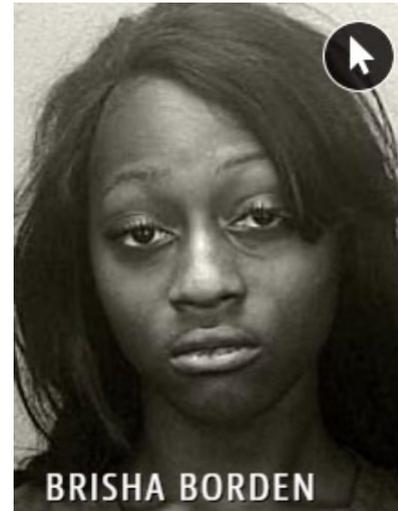
Früher Straftaten:

4 Jugendvergehen

Risk Assessment Solution/Output

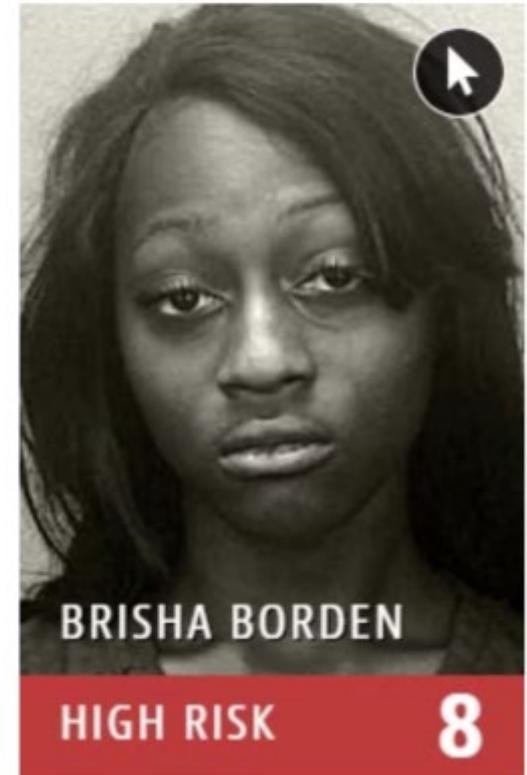
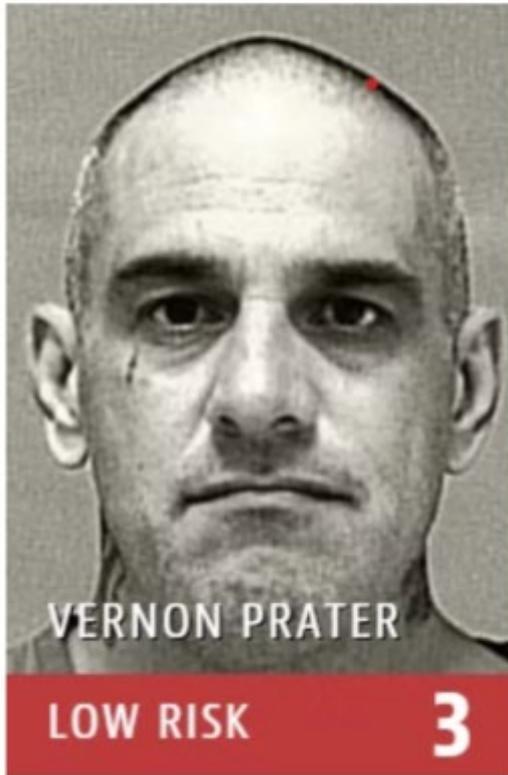


Errechneter Risikoscore
durch COMPAS?

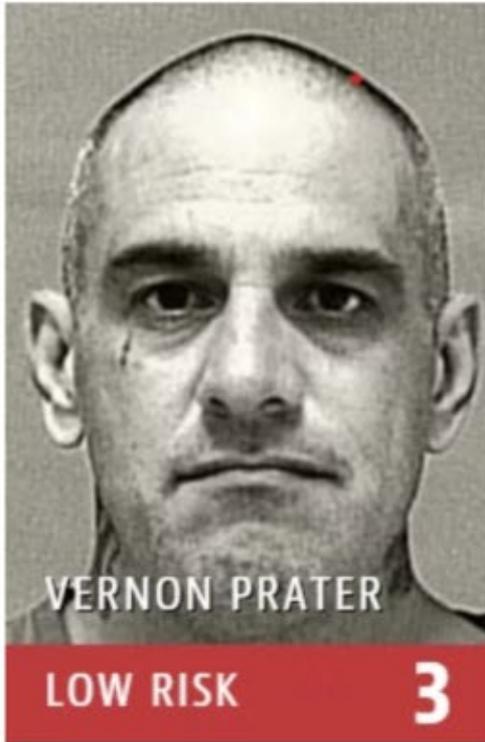


Errechneter Risikoscore
Durch COMPAS?

Risk Assessment Solution



Risk Assessment Solution



Subsequent Offences:

1 grand theft



Subsequent Offences:

None

Every line of code is a decision



Security-Konferenz München 2022



Pic von Frau mit Tattoos



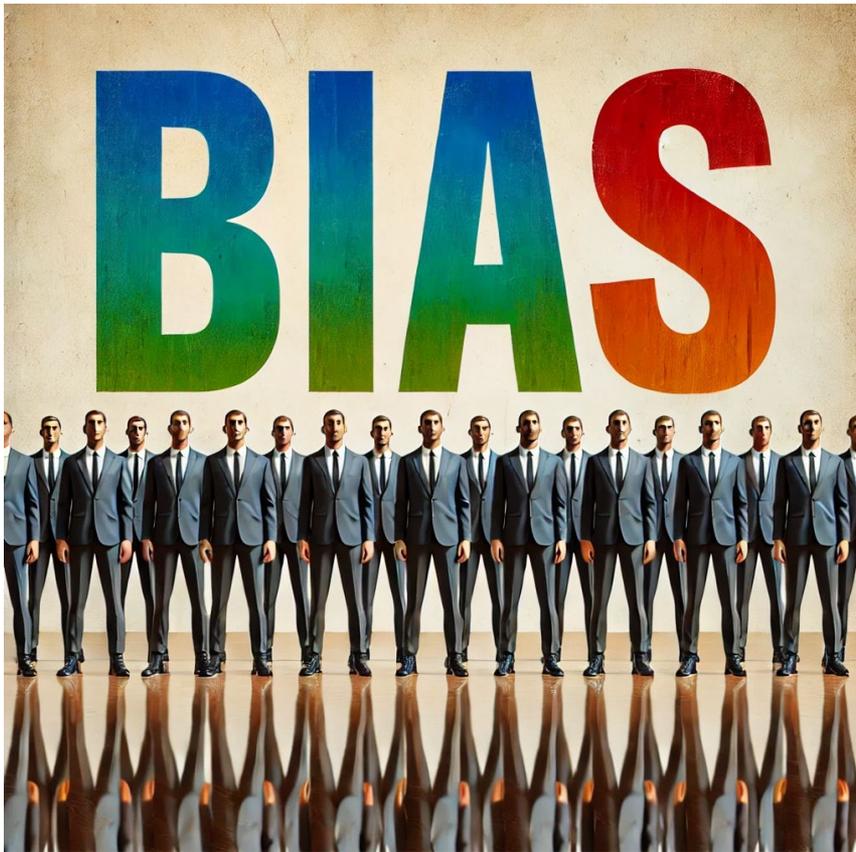
Canva-integriertes Text-zu-Bild KI
Magic Media

Prompt:
„Erstelle mir ein Bild einer tätowierten
Frau“

Terms & Conditions:

Sicherheit der Plattform ist zu gewähren
Keine sexuell eindeutigen Inhalte
Keine Themen zu Selbstverletzung

Thomaskreislauf



1. **Fehler** aus historischen Daten
2. subjektive Fehleinschätzung bei Gestaltung von Algorithmen
3. **Feedback-Loops**, die die Verzerrungen verstärken.
4. **systematische** Diskriminierung
5. **Soziale** Akzeptanz von KI-Vorhersagen (Automation Bias, Blindspot Bias)

Was ist Bias und wie entsteht er?

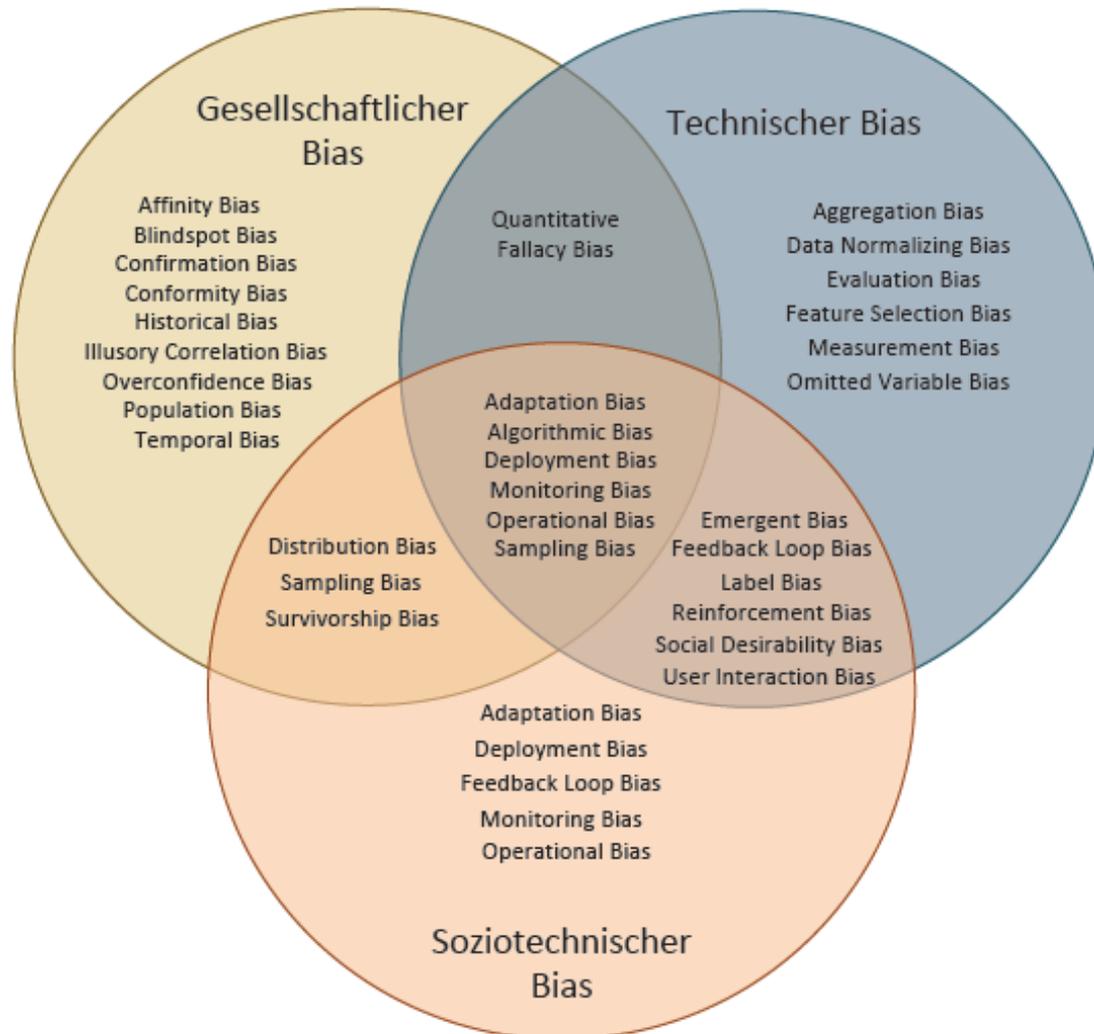
KI-Bias tritt auf, wenn ein Algorithmus als Folge **fehlerhafter Annahmen** im ML-Prozess **systematisch verzerrte Ergebnisse** liefert und somit **bestimmte Gruppen ungerecht bevorzugt oder benachteiligt**.

Technische Fehler
Falschmessung
Sensorische Fehler
STATISTIK

Systematische Verzerrung
Denken, Vorurteil
KOGNITIONSPSYCHOLOGIE



32 identifizierte Bias-Arten bei KI-System





ChatGPT:

Generiere mir ein **Bild**
von einer **Frau** mit
Herzinfarkt-
Symptomen.

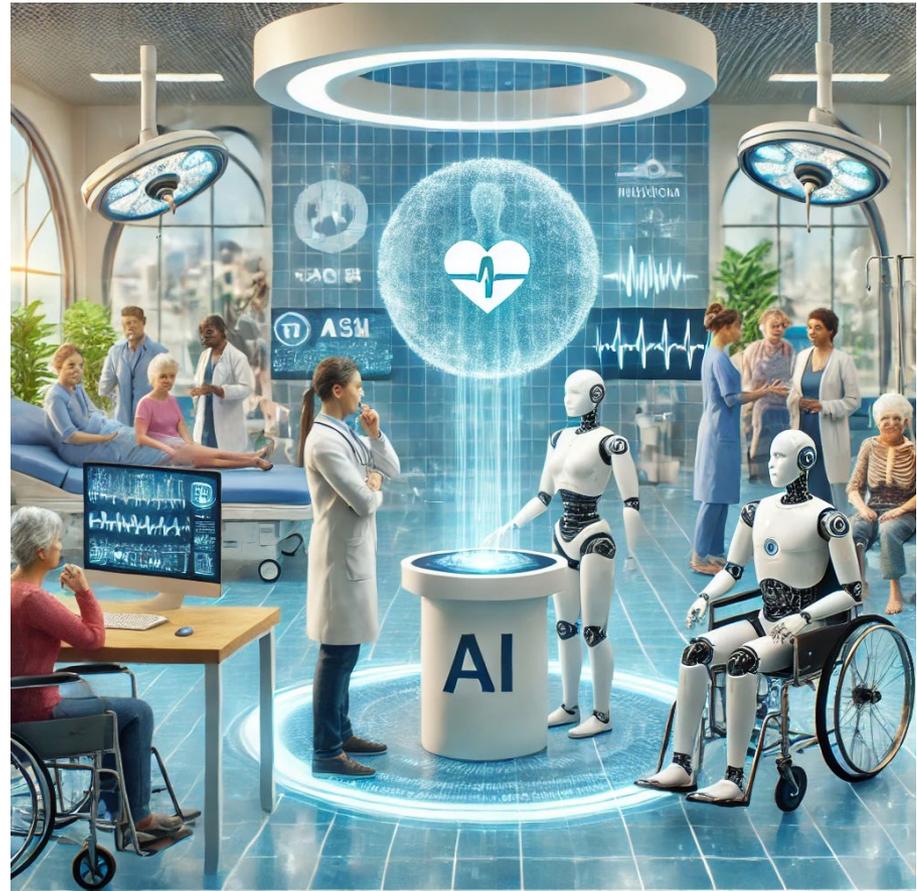


Mit einer **korrekten & repräsentativen** Datenlage würde das Bild so aussehen

KI für mehr Inklusion

Gesundheitsversorgung

- Inklusive Gesundheitsforschung (Gender Medizin, kuratierte Daten, barrierefreie Datenerhebung)
- Personalisierte Therapien & Präzisions-Medizin
- Telemedizin, virtuelle Assistenten, Frühwarnsysteme
- Früherkennung & Prävention von Krankheiten (Wearables)
- Patienten-Empowerment (Gesundheitsaufklärung –Info in unterschiedlichen Sprachen/Formaten)



KI für mehr Inklusion



- Barrierefreie Kommunikation
(Sprachübersetzung, Text to Speech...)
- Gesellschaftliche Teilhabe (virtuelle Events, Museumsbesuche)
- Assistive Technologien (Smart Home Mobilität, autonomes Fahren)
- E-Government & Chatbots
- Bildungszugang (Personalisiertes Lernen, inklusive Lehrinhalte, Bildungsressourcen)
- Citizen Science (Crowdsourcing)

Dual Use – Fluch oder Segen?



Agenda – Wozu genützt?

Wirtschaftlicher Gewinn & **Werbeeinnahmen** (Clickbait..)!!

Politische **Einflussnahme** & Destabilisierung der
Demokratie

Propaganda & ideologische Verbreitung (wenig Info, viel
Meinung)

Psychologische **Manipulation** (Think Tanks – Blick ins
Impressum)

Schüren von Angst, Verwirrung, Panik, Misstrauen

Cyberkriminalität

Phishing, Social Engineering, Malware, Identitätsdiebstahl, Fake Shops...



KI aber auch für Fake News/Manipulation

Thousands go to fake AI-invented Dublin Halloween parade



November 2024 Dublin
Gruselige Halloween News

Sachverhalt:

- pakistanische Website erstellte News-Artikel für Werbeeinnahmen und erfand einen Fake Event in Dublin
- Durch SEO gelangte dies in die Mainstream-Medien
- Tausende erschienen zu diesem nicht-existierenden Event

Problematisch:

- Menschen konsumieren offenbar KI-generierte Fake-News in einem für Werbeeinnahmen profitablen Ausmaß.
- Mainstream-Medien versäumten grundlegende Quellenprüfung.
- Tausende überprüften die Quelle nicht, bevor sie das Event besuchten.

Mother says AI chatbot led her son to kill himself in lawsuit against its maker

Megan Garcia said Sewell, 14, used Character.ai obsessively before his death and alleges negligence and wrongful death



Oktober 2024 Florida

**Chatbot Dany: „Come home to me“
Sewell: “I promise I will come home to you. I love you so much, Dany.”**

- Mutter verklagt Character.ai wegen Suizid ihres 14-jährigen Sohnes nach intensiver Nutzung eines KI-Chatbots
- Chatbot führte laut Klage Jugendliche in die Depression und bestärkte Suizidgedanken
- „Gefährliches KI-Produkt“: Familie fordert Rechenschaft von Character.ai und Google

AI Act – ein Rechtsrahmen für KI in der EU



- Ab **1. August 2024** in Kraft
- **Risikobasierter** Ansatz
- Erste weltweite KI-Regulierung
- mit **rechtlich & ethischer** Relevanz!

Intention des gemeinsamen Vorgehens:

- Erschließen des Potenzials von KI-Technologien
- Minimieren von Risiken
- Förderung von Chancen mit KI
- Vermeidung von Diskriminierungen
- Transparente Nutzung

AI Act – 4 Risikoklassen

ARTIFICIAL INTELLIGENCE ACT



Quelle: digital-strategy.ec.europa.eu

© Quelle: digital-strategy.ec.europa.eu

MARTINA GAISCH

ISABELLA MADER

NEXT GENERATION DIGITAL

AI Ethics & Human Factors in AI



AI Ethics & Human Factors in AI

Mehr als ein Buch. Ein Leitfaden für Menschen mit Verantwortung in einer modernen, technologiegetriebenen Gesellschaft.

In einer Welt, die zunehmend von künstlicher Intelligenz durchdrungen ist, stehen wir vor einer entscheidenden Frage: Wie können wir sicherstellen, dass KI-Systeme nicht nur effektiv, sondern auch ethisch und verantwortungsvoll sind?

„AI Ethics & Human Factors“ beleuchtet die Bedeutung des menschlichen Einflusses in einer von Algorithmen dominierten Welt und zeigt, wie Verantwortung angemessen und vertrauensstiftend wahrgenommen werden kann.

Zum Inhalt

- KI SICHTET: KI erkennt Muster als Black Box – die wir verstehen lernen müssen.
- KI DICHTET: Die kreative Seite der KI und damit verbundene Herausforderungen.
- KI RICHTET: Ethische und rechtliche Implikationen von Entscheidungen einer KI.
- KI VERNICHTET: Die Risiken und destruktiven Potenziale von KI.
- DER MENSCH DENKT: Menschliches vs. maschinelles Lernen, und Lernen voneinander!
- DER MENSCH ERKENNT: Bias in KI-Systemen und in uns selbst erkennen
- DER MENSCH BEDENKT: Richtlinien und rechtliche Rahmenbedingungen für ethische KI.
- DER MENSCH LENKT: Empfehlungen für faire und inklusive KI-Systeme.

Dieses Buch ist: Ein Aufruf zum Handeln in einer digitalen Ära.

FH-Prof.ⁱⁿ Mag.^a Dr.ⁱⁿ Martina Gaisch ist promovierte Bildungssoziologin, Hochschulforscherin und Professorin an der Fachhochschule OÖ, Fakultät für Informatik, Kommunikation und Medien in Hagenberg. Sie lehrt unter anderem Technikfolgenabschätzung, Ethics & Responsible AI.

Isabella Mader BA MMSc ist Vorstand von Excellence Research. Sie wurde bereits als Top CIO des Jahres ausgezeichnet. An der FH Hagenberg lehrt sie seit 15 Jahren IT- und KI-Strategie, Human Factors in AI und begleitet Studienprojekte mit KILM-Schwerpunkten.

METHODEN



REGULAREN



KONZEPTE



SZENARIEN



CASES



Excellence Edition



9 783991 394334



KI & ME DER KI VIDEOCAST

<https://www.youtube.com/watch?v=tsQa1CRDleI>

Gutes Gelingen und vielen Dank!

FH-Prof.ⁱⁿ Mag.^a Dr.ⁱⁿ Martina Gaisch

martina.gaisch@fh-hagenberg.at

Vernetzen wir uns auf LinkedIn



Studiengangsleitung

„Design of Digital Products“ an der FH OÖ

Fakultät für Informatik, Kommunikation & Medien Hagenberg

<https://www.fh-ooe.at/campus-hagenberg/studiengaenge/bachelor/design-of-digital-products/>

