

Strategiebriefing Taskforce „KI in der Hochschulbildung“ Stifterverband

**Bias & KI –
Sensibilisierung für Bias in KI-Systemen, kritische Reflexion und
Diskussion über Handlungsmöglichkeiten im Hochschulkontext**

Antonia Freesemann, M.A.
Dresden, 14.10.2025

Überblick

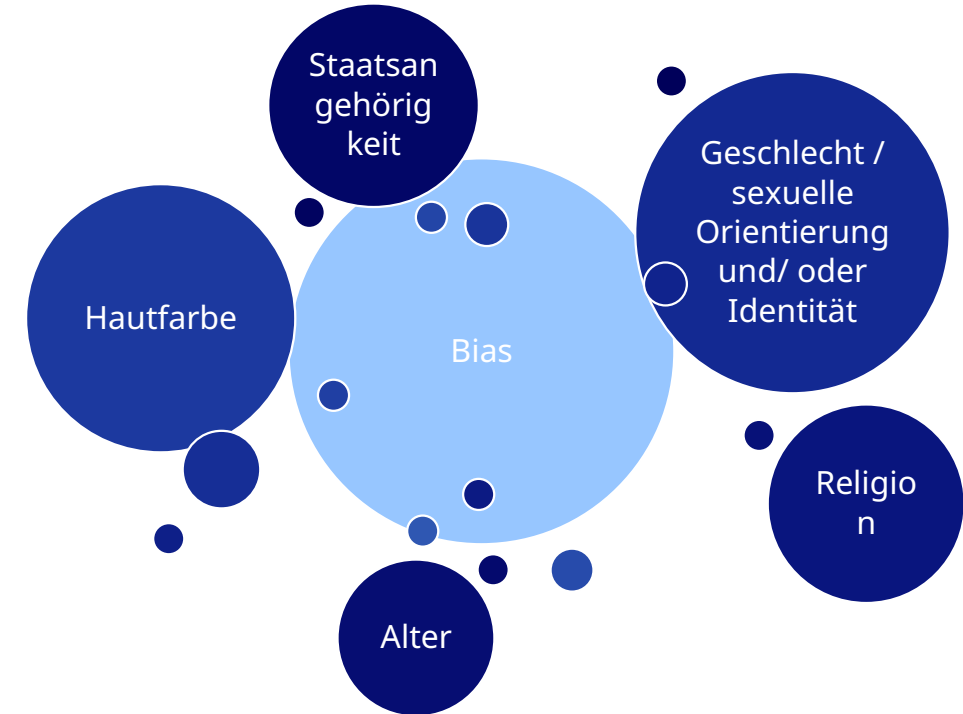
Einleitung: Definition,
Formen & Ursprung

Bias in KI-Systemen

Ansätze &
Handlungsmöglichkeiten
im Hochschulkontext

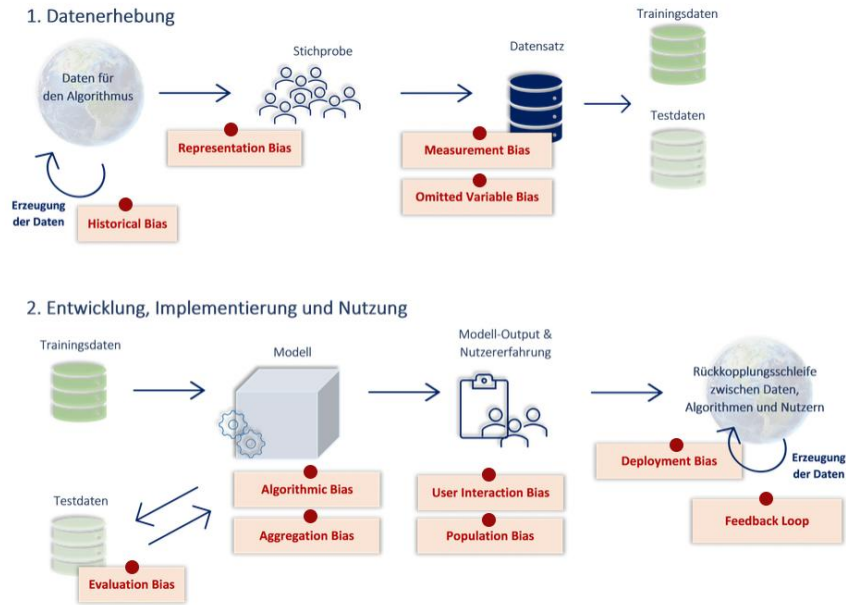
Einführung

- Ausgangspunkt & Kontext: Themen abseits von Fragen zu technischen Möglichkeiten und Limitationen der Technologie treten in den Fokus
- Zentrale Problemstellung: Die Ungleichbehandlung [...] innerhalb des Datenraumes durch KI-Systeme (Angwin, et al., 2016) (Buolamwini, et al., 2018) (Dastin, 2022) (ISO/IEC TR 24027, 2021)
- Fehlannahme der Objektivität von Technik
 - Realität: Vermeintlich neutrale Instanz der Technik kann Diskriminierungseffekte durch geringe Nachvollziehbarkeit entstehen lassen und bestätigen
- Was bedeutet (algorithmischer/ technischer) Bias? = Verzerrungen, Ungleichbehandlung von KI-Systemen

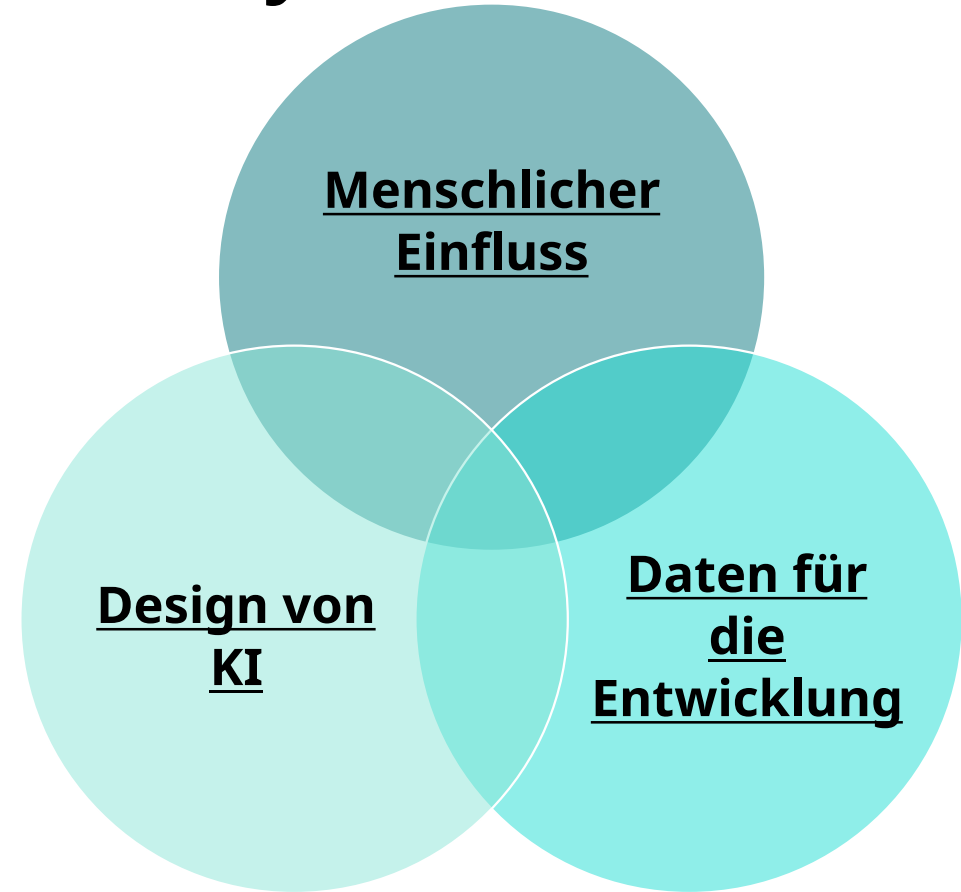


Vgl. Bundesamt für Sicherheit in der Informationstechnik (2025)

Formen & Gründe von Bias in KI-Systemen



Vgl. Institut für Business Analytics der Universität Ulm (o.J.)



Vgl. Gengler, E., Kraus, A., Bodrožić-Brnić, K. (2024)

Formen & Gründe von Bias in KI-Systemen

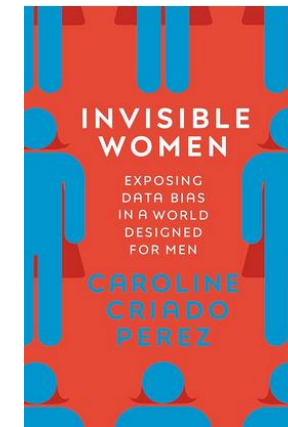
Globale Verteilung KI-Branche:
22% Frauen / 78% Männer
Deutschlandweite Verteilung KI-Branche:
20% Frauen / 80% Männer

Vgl. Pal S.; Marino Lazzaroni, P. & Mendoza, P. (2024)

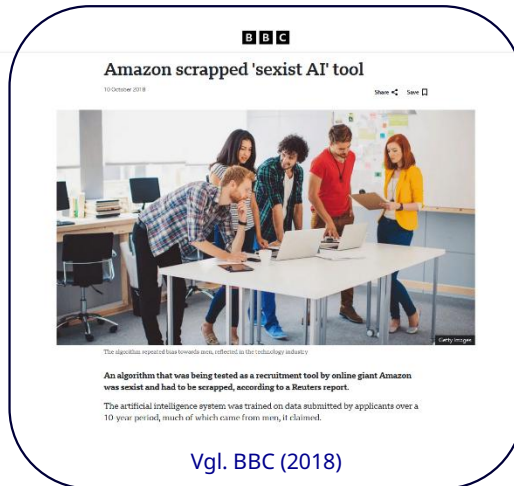
Menschlicher
Einfluss

Design von
KI

Daten für
die
Entwicklung



Criado Perez (2019)



Vgl. BBC (2018)

Vgl. Gengler, E., Kraus, A., Bodrožić-Brnić, K. (2024)

→ Bias entsteht nicht nur durch verzerrte Daten. Bias kann auch aus der Art und Weise resultieren, wie das KI-System modelliert oder evaluiert wird bzw. wie Nutzer:innen die Endergebnisse des KI-Systems interpretieren

Beispiele von Bias in KI-Systemen

Autonomes Fahren

Schau mir in die Augen, Kleinwagen!

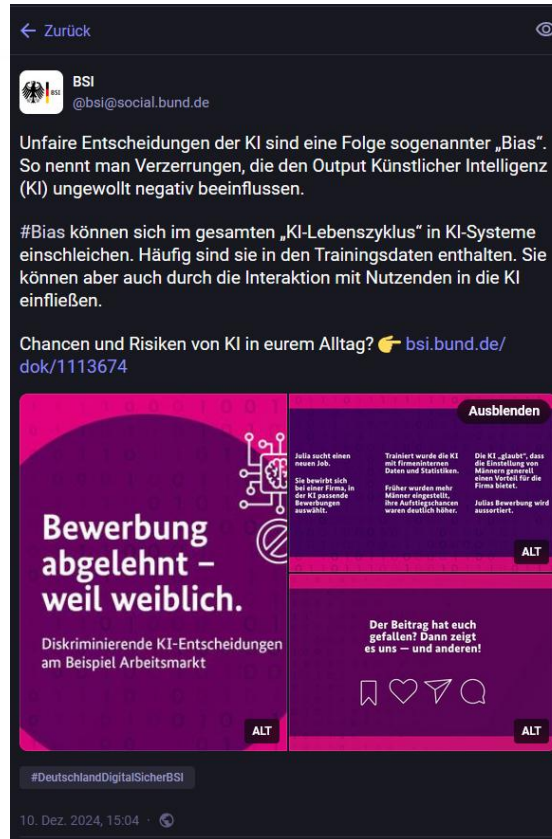
In Japan hat ein autonom fahrender Bus einen Mann mit Sehbehinderung angefahren, weil die Systeme sein Verhalten nicht richtig vorhersagen konnten. Immer wieder werden die Bedürfnisse von Menschen mit Behinderung bei der Technologieentwicklung vernachlässigt. Damit kämpfen auch Entwickler:innen in Deutschland.

04.10.2021 um 07:10 Uhr - Jana Ballweber - in Technologie - 23 Ergänzungen



Autonome Fahrzeuge sollen irgendwann Blickkontakt mit Passant:innen aufnehmen. Das gefährdet

Vgl. Ballweber (2021)



Vgl. Beitrag BSI bei X (2024)

BACKLASH

Vorurteile und zweifelhafte Umsetzung: AMS-KI-Chatbot trifft auf Spott und Hohn

Der auf ChatGPT basierende "Berufsinformat" zeigt allerlei Probleme – und lässt sich zudem leicht austricksen. Das AMS weist die Kritik von sich, sieht Kosten von 300.000 Euro gerechtfertigt

Andreas Proschofsky

aktualisiert am 4. Jänner 2024, 13:03

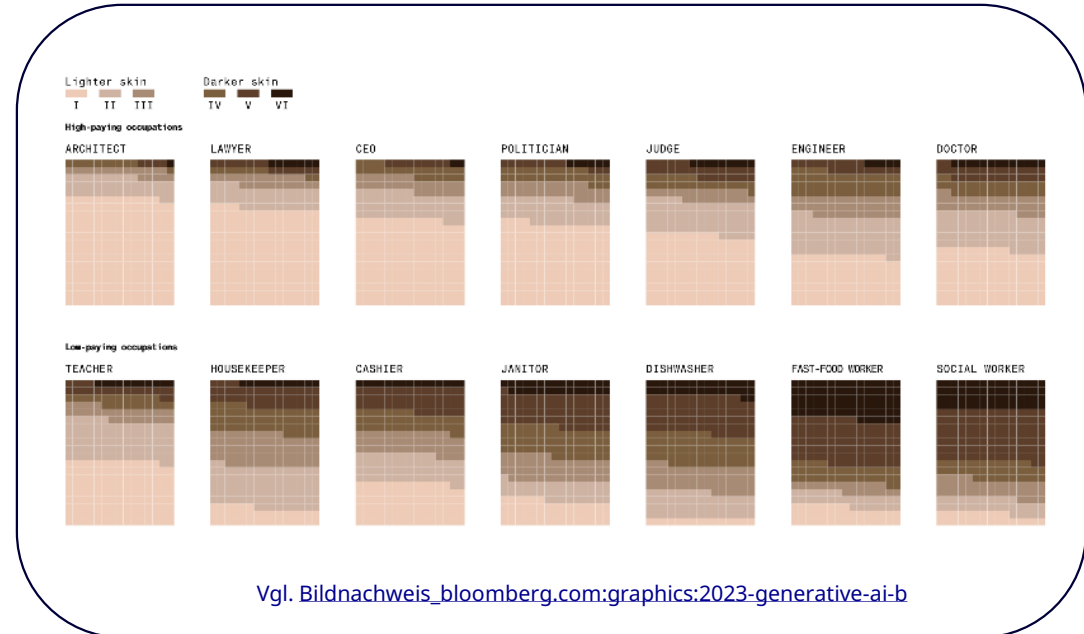
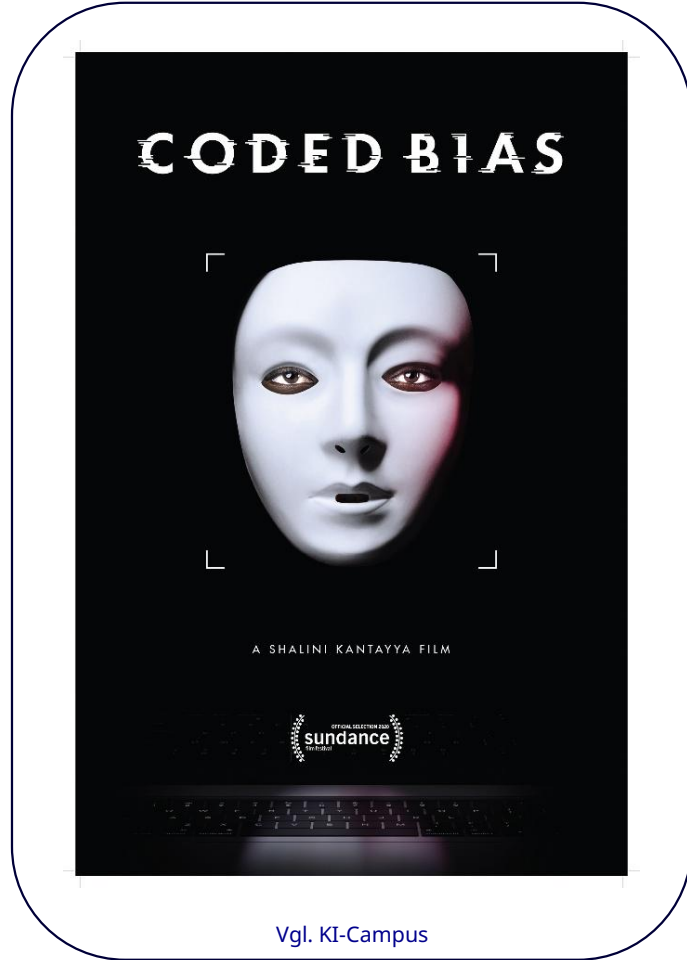
1121 Postings Später lesen

Die Idee klingt zumindest nicht ganz verkehrt. Mit dem Berufsinformat hat das Arbeitsmarktservice (AMS) am Mittwoch einen eigenen KI-Chatbot vorgestellt, der Arbeitssuchenden bei Fragen zu Jobmöglichkeiten hilfreich zur Seite stehen soll. Wie sich in den Stunden danach schnell herausstellen sollte, macht der Service dies auch und mag in manchen Teilen deshalb unzweifelhaft nützlich sein. Gleichzeitig ist aber auch schnell massive Kritik an der Umsetzung aufgekommen. Denn wie sich herausstellt, gibt es nicht nur offenbar eklatante Mängel bei der technischen Implementation des Dienstes – das Tool produziert auch eifrig Vorurteile.



Vgl. Proschofsky (2024)

Beispiele von Bias in KI-Systemen



Beispiele von Bias in KI-Systemen

10. 06. 2025

Verfasst von: Ekaterina Lapshinova-Koltunski, Maja Popović

Gender-Bias in KI-übersetzten Amazon-Rezensionen



Beim Onlinekaufen orientieren wir uns an Produktrezensionen. Hat es dabei einen Einfluss, ob die Bewertung von einer Frau oder einem Mann stammt? Werden die Kommentare von einer KI in unsere Sprache übersetzt, dominieren Worte in männlicher Form.

Vgl. Lapshinova-Koltunski & Popović (2025)

Menü

DIE ZEIT

Abo testen

Apple Card

Weiblich, Ehefrau, kreditunwürdig?

Die Apple Card soll Frauen weniger Kredit gewährt haben als Männern. Gelebte Bankenpraxis. Sich gegen Algorithmen zu wehren ist aber noch schwerer als gegen Bankberater.

Eine Analyse von [Lisa Hegemann](#)

21. November 2019, 15:56 Uhr

283

Inhalt

Auf einer Seite lesen

David Heinemeier Hansson muss ziemlich wütend gewesen sein auf Apple. Genauer gesagt: auf die Apple Card, die Kreditkarte des Unternehmens, die seit August in den USA verfügbar ist. Und wie man das eben so macht als einflussreicher Softwareentwickler mit gut 360.000 Twitter-Followern, tat er diesen Ärger von einigen Tagen auf Twitter kund.

Die Apple Card sei "verdammt sexistisch", schrieb er in einer Serie von Tweets. Seine Frau Jamie und er gäben gemeinsame Steuererklärungen ab und lebten in einer Errungenschaftsgemeinschaft - dennoch habe er einen 20-mal höheren Kreditrahmen als seine Partnerin. Dabei habe sie sogar eine höhere Bonität, wird also als kreditwürdiger eingestuft als er.

Vgl. Hegemann (2019)

Beispiele von Bias in KI-Systemen

LARGE LANGUAGE MODELS

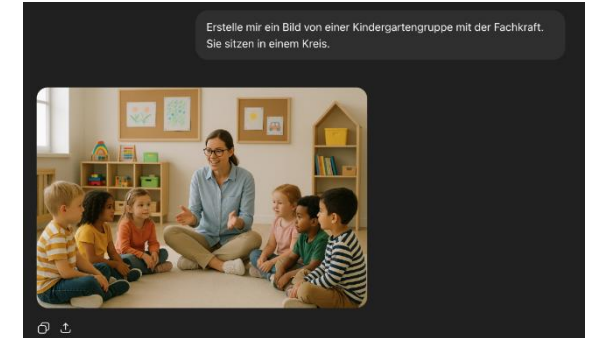
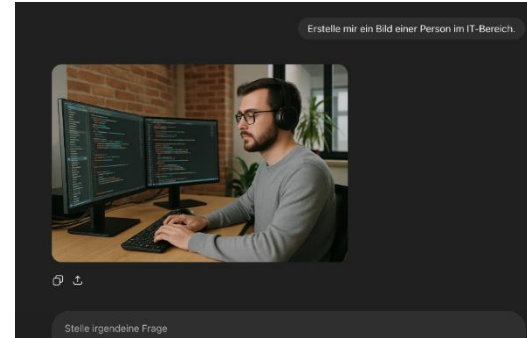
Weniger faul und weniger fleißig: Wie KI Vorurteile über Ostdeutsche kreiert

13. Oktober 2025, 09:41 Uhr

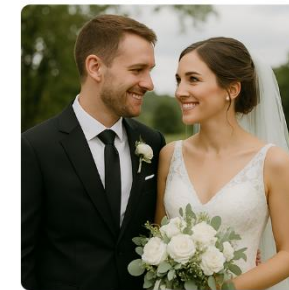
Künstliche Intelligenz ist nicht frei von Vorurteilen und falschen Einschätzungen. Auch nicht, wenn es um Ostdeutsche geht, wie eine Untersuchung aus München zeigt.

von MDR WISSEN

Vgl.: MDR WISSEN (2025)



Create now a image of a couple who is about to get married.



Quelle: ChatGPT (GPT-4o)

Einsatz KI-Systeme im Hochschulkontext (vgl. Gengler 2025 / Mosene & Leitfeld 2025):

Lehre

- Generative KI in Lehre (AI Literacy)
- Intelligente Tutoring-Systeme
 - Learning Analytics
 - Personalisierte Empfehlungen
 - Evaluation
 - Sprach- und Übersetzungstools

Verwaltung

- Chatbots
- Prozessautomatisierung
- Kapazitäts- & Raumplanung
- Bewerbungsmanagement
 - Betrugsprävention

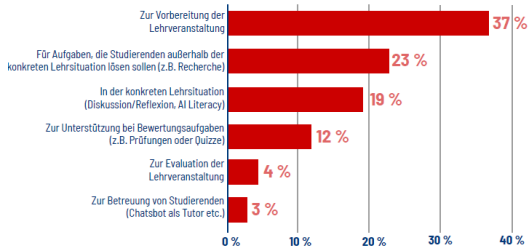
Forschung

- Citizen Science & Human-in-the-Loop-Modelle
- Datenanalyse & Mustererkennung
- Literaturrecherche & Wissensmanagement
- Textgenerierung & Zusammenfassung
 - Peer-Review & Qualitätssicherung
- Forschungsmanagement

Studierende

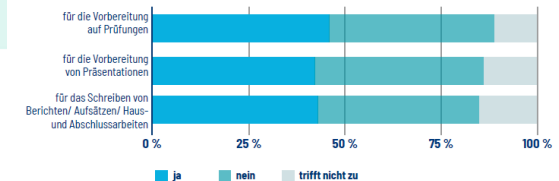
- Schreibassistenz
- Präsentationen/ Bilder/ Videos erstellen
- Lernmanagement
- Programmierfähigkeit
- Feedbacksysteme

Wofür KI verwendet wird:



Lehrende: Für welche der nachfolgenden Prozesse verwenden Sie KI-Anwendungen? Mehrfachnennung möglich | n = 500 | Angabe in Prozent | Keine Angabe von Sonstiges

Zu welchen Zwecken nutzen Studierende künstliche Intelligenz in ihrem Studium?



Nutzen Sie künstliche Intelligenz (wie z.B. ChatGPT) in Ihrem Studium zu folgenden Zwecken? | n = 811 | Angabe in Prozent | keine Angabe von sonstiges | Auswahl von Antwortoptionen

Vgl. Budde, Tobor & Friedrich (2024)



Vgl. Budde, Tobor & Friedrich (2024)



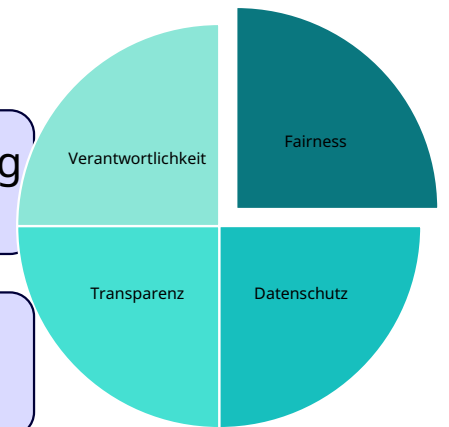
Fairness im Kontext Künstlicher Intelligenz

Fairness = Alle Menschen werden gleichberechtigt und diskriminierungsfrei von KI-Systemen behandelt

- ✓ Gleiche Chancen
- ✓ Verbot der Diskriminierung
- ✓ Transparenz und Nachvollziehbarkeit
- ✓ Rechenschaftspflicht

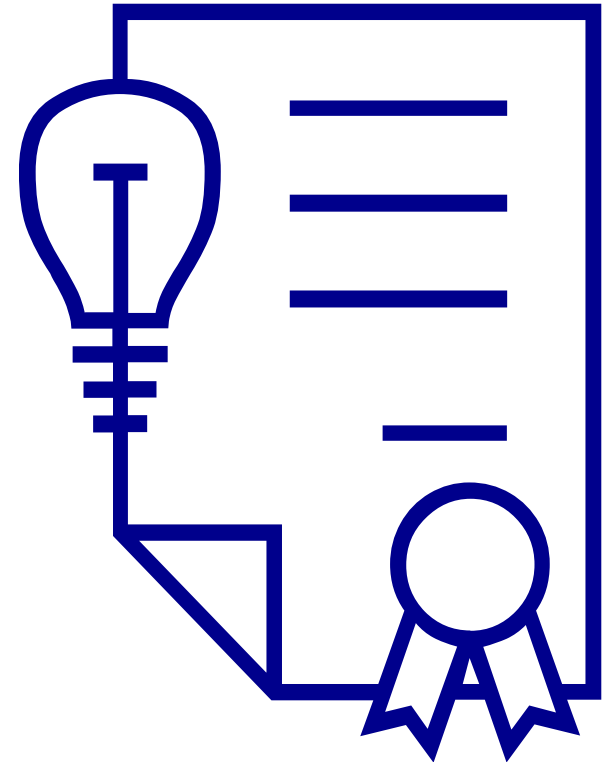
→ Fairness ist eine wichtige, aber nicht einzige Prämisse, die bei der Entwicklung und Nutzung von KI berücksichtigt werden sollte

→ **KI-Ethik**



Ideen & Empfehlungen: Strategische Verankerung & Governance

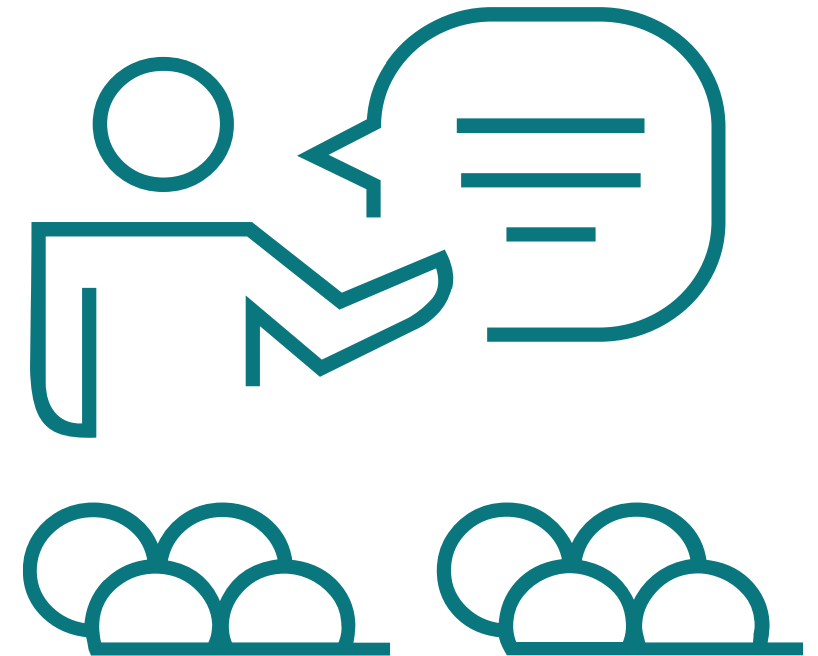
- Integration in (KI-)Hochschulstrategie
 - Festlegung ethischer Grundlagen (z.B. Transparenz, Nachvollziehbarkeit, Rechenschaft, Nichtdiskriminierung)
 - Regelmäßige Evaluierungsmechanismen
- Langfristige Vision (z.B. nach TUM-Vorbild)
 - Stärkung des Vertrauens in KI-Systeme
 - Förderung eines integrativen akademischen Umfelds
 - Benchmark-Institution
- Ethische KI-Leitlinien & Handreichungen
 - Bias-Prüfung
 - Algorithmische Audit-Verfahren
 - Ethikfreigabeprozesse
 - Transparenzberichte
- Unterstützungsinstrumente z.B. Implementierungsrahmen



Vgl. Braun (2024) TUM KI Strategie

Ideen & Empfehlungen: Lehre & Studium

- Curriculare Verankerung & didaktische Maßnahmen
 - Pflichtmodule
 - Wahlpflichtmodule
 - Praxisorientierte Lehre
- Weiterbildung Lehrende
 - Sensibilisierung & Maßnahmen zur sicheren Nutzung von KI-Anwendungen in der Lehre z.B. Bestandsaufnahme, Risikoeinschätzung, menschliche Letztkontrolle
- Support-Strukturen
 - Feedback-Mechanismen
 - Beratung



Ideen & Empfehlungen: Forschung & Transfer

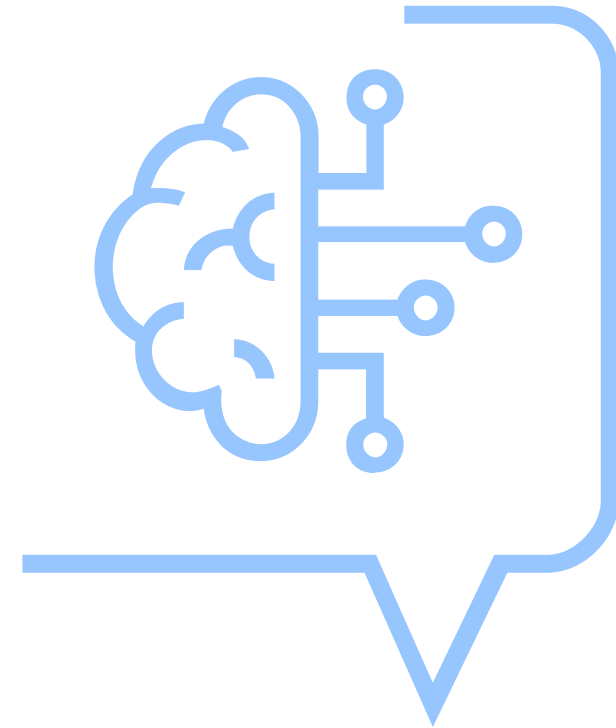
- Bias-fokussierte Forschungsförderung
 - Bias Detection Tools
- KI-Systeme entwickeln & Integration von Fairness sicherstellen
 - Verantwortungsbewusste KI-Entwicklung
 - Interdisziplinäre Teams
 - Transferorientierung
- Offene Wissenschaft & gesellschaftlicher Dialog
 - Open-Source-Tools



Vgl. Braun (2024) TUM KI Strategie

Ideen & Empfehlungen: Kultur & Kommunikation

- Kultur des Vertrauens & der Verantwortung (z.B. nach TUM-Vorbild)
 - Offene Diskursräume
 - Weiterbildungsprogramme für alle Statusgruppen
 - Anerkennung bias-sensibler Initiativen
- Awareness-Kampagnen & hochschulweite Initiativen
- Kommunikation & Transparenz
 - KI-Einsatz, Werte, Prinzipien
 - Transparente Darstellung: Projekte, Leitlinien und Initiativen
 - Communities of Practice
- Dialog & Partizipation
- Anreiz- & Belohnungssysteme



Vgl. Braun (2024) TUM KI Strategie

Vielen Dank für Ihre Aufmerksamkeit!

Literatur- und Quellenverzeichnis

Ballweber, J. (2021): Autonomes Fahren. Schau mir in die Augen, Kleinwagen! URL: <https://netzpolitik.org/2021/autonomes-fahren-schau-mir-in-die-augen-kleinwagen/> , Zuletzt geöffnet: 08.08.2025

BBC (2018): Amazon scrapped "sexist AI" tool. URL: <https://www.bbc.com/news/technology-45809919> , Zuletzt geöffnet: 08.08.2025

BSI (2024): Beitrag auf X. URL: <https://social.bund.de/@bsi/113628902274356300> , Zuletzt geöffnet: 08.08.2025

Budde, J., Tobor, J., Friedrich J. (2024). Künstliche Intelligenz. Wo stehen die deutschen Hochschulen?. Berlin: Hochschulforum Digitalisierung.

Bundesamt für Sicherheit in der Informationstechnik (2025): Bias in der künstlichen Intelligenz. Bonn: Bundesamt für Sicherheit in der Informationstechnik.

Braun, A. (2024): TUM KI Strategie.

ChatGPT version -4o: Hilfe bei der Zusammenfassung und Erstellung der Textstruktur. OpenAI: openai.com/chat.

Criado Perez, C. (2019): Invisible Women: Exposing Data Bias in a World Designed for Men. Chatto & Windus.

Fraunhofer IKS (o.J.): Künstliche Intelligenz (KI) und maschinelles Lernen. URL: <https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html> , Zuletzt geöffnet: 07.08.2025

Gengler, E. (2025): Macht & Bias in KI: Feministische Antworten in der Hochschule. U:FF. URL: <https://www.youtube.com/watch?v=NunzbF1sjvk> , Zuletzt geöffnet: 08.08.2025

Gengler, E., Kraus, A., Bodrožić-Brnić, K. (2024): Faires KI-Prompting – Ein Leitfaden für Unternehmen. BSP Business and Law School – Hochschule für Management und Recht. (1 – 28).

Gössl, S. & Yakar, S. (2023): Geschlechtsneutrale KI. Eine Handreichung. Land Schleswig-Holstein: Ministerium für Soziales, Jugend, Familie, Senioren, Integration und Gleichstellung.

Hegemann, L. (2019): Weiblich, Ehefrau, kreditunwürdig? Die Apple Card soll Frauen weniger Kredit gewährt haben als Männern. Gelebte Bankenpraxis. Sich gegen Algorithmen zu wehren ist aber noch schwerer als gegen Bankberater. URL: <https://www.zeit.de/digital/datenschutz/2019-11/apple-card-kreditvergabe-diskriminierung-frauen-algorithmen-goldman-sachs> , Zuletzt geöffnet: 08.08.2025

Literatur- und Quellenverzeichnis

Hochschule Bayern (o.J.): KI-Leitlinie Hochschullehre. Empfehlungen zum Umgang mit KI in der Lehre der bayrischen Hochschulen für angewandte Wissenschaften und Technischen Hochschulen. URL: https://www.hochschule-bayern.de/fileadmin/daten/Positionspapiere/KI-Leitlinie_Hochschule-Bayern-19052025.pdf , Zuletzt geöffnet: 08.08.2025

Hüsch, M.; Horstmann, N. & Breiter, A. (2025): Künstliche Intelligenz im Studium – die Sicht von Studierenden im Wintersemester 2024/25. URL: <https://hochschuldaten.che.de/kuenstliche-intelligenz-im-studium-die-sicht-von-studierenden-im-wintersemester-2024-25/> , Zuletzt geöffnet: 08.08.2025
Institut für Business Analytics der Universität Ulm (o.J.): Grundlagen - Bias und Fairness in KI. Bias und Fairness in KI. <https://bias-and-fairnessin-ai-systems.de/grundlagen/>

Kapoor, S. & Narayanan, A. (2023): Quantifying ChatGPT's gender bias. Benchmarks allow us to dig deeper into what causes biases and what can be done about it. URL: <https://www.aisnakeoil.com/p/quantifying-chatgpts-gender-bias> , Zuletzt geöffnet: 08.08.2025

KI-Campus (o.J.): Können Maschinen diskriminieren? URL: <https://ki-campus.org/texts/diskriminierung> , Zuletzt geöffnet: 08.08.2025

Lapshinova-Koltunski, E. & Popović, M. (2025): Gender-Bias in KI-übersetzten Amazon-Rezensionen. URL: <https://www.wissenhochn.de/de/themen/auswahl-und-uebersicht/einzelsicht/gender-bias-in-ki-uebersetzten-amazon-rezensionen> , Zuletzt geöffnet: 08.08.2025

MDR Wissen (2025): Large Language Models Weniger faul und weniger fleißig: Wie KI Vorurteile über Ostdeutsche kreiert. URL: <https://www.mdr.de/wissen/naturwissenschaften-technik/weniger-faul-und-weniger-fleissig-llms-kuenstliche-intelligenz-diskriminieren-ostdeutsche-100.html> , Zuletzt geöffnet: 13.10.2025.

Mittelstand-Digital. Zentrum Fokus Mensch (2023). URL: <https://www.digitalzentrum-fokus-mensch.de/kos/WNetz?art=News.show&id=2164> , Zuletzt geöffnet: 08.08.2025 / [Bildnachweis_bloomberg.com/graphics/2023-generative-ai-b](https://www.bildnachweis.com/graphics/2023-generative-ai-b)

Mosene, K. & Leitfeld, J. (2025): Bias erkennen, Verantwortung übernehmen: Kritische Perspektiven auf KI und Datenqualität in der Hochschulbildung. URL: <https://hochschulforumdigitalisierung.de/bias-erkennen-kritische-perspektiven-auf-ki/> , Zuletzt geöffnet: 13.10.2025.

Pal, S.; Marino Lazzaroni, R. & Mendoza, P. (2024): AI's Missing Link: The Gender Gap in the Talent Pool. Interface. URL: https://www.interface-eu.org/publications/ai-gender-gap?utm_campaign=DE%20Community%20NL%20Weltfrauentag&utm_medium=email&utm_source=Mailjet , Zuletzt geöffnet: 08.08.2025

Proschofsky, A. (2024): Vorurteile und zweifelhafte Umsetzung: AM-KI-Chatbot trifft auf Spott und Hohn. URL: <https://www.derstandard.at/story/3000000201774/vorurteile-und-zweifelhafte-umsetzung-der-ams-ki-chatbot-trifft-auf-spott-und-hohn> , Zuletzt geöffnet: 08.08.2025

West, S.M., Whittaker, M. and Crawford, K. (2019): Discriminating Systems: Gender, Race and Power in AI. AI Now Institute.