

Infotext für Schülerinnen und Schüler

So entsteht Bias in KI

Künstliche Intelligenz (KI) hat das Potenzial, viele Probleme unseres Alltags zu lösen, bringt jedoch zeitgleich Herausforderungen mit sich, insbesondere durch den sogenannten „Bias“. Bias bezeichnet einen Verzerrungseffekt, der dazu führen kann, dass KI-Systeme unfaire oder diskriminierende Entscheidungen treffen.

Trainingsdaten

Bias in KI entsteht oft durch die Trainingsdaten. Wenn die Daten, mit denen eine KI trainiert wird, nicht vielfältig genug sind oder bestimmte Gruppen über- oder unterrepräsentiert sind, kann die KI verzerrte Muster lernen. Beispielsweise kann eine Bilderkennungs-KI, die hauptsächlich mit Bildern von hellhäutigen Menschen trainiert wurde, Schwierigkeiten haben, dunkelhäutige Menschen korrekt zu erkennen.

Algorithmen (Überbewertung)

Algorithmen können ebenfalls zu Bias beitragen, indem sie bestimmte Merkmale überbewerten. Wenn ein Algorithmus zum Beispiel lernt, dass eine bestimmte Eigenschaft stark mit einem Ergebnis korreliert, kann er dieser Eigenschaft zu viel Gewicht beimessen und andere wichtige Faktoren vernachlässigen.

Menschliche Voreingenommenheit der Entwicklerinnen und Entwickler

Die Entwickler und Entwicklerinnen, die KI-Systeme programmieren, bringen ihre eigenen unbewussten Vorurteile mit ein. Diese menschlichen Bias können sich

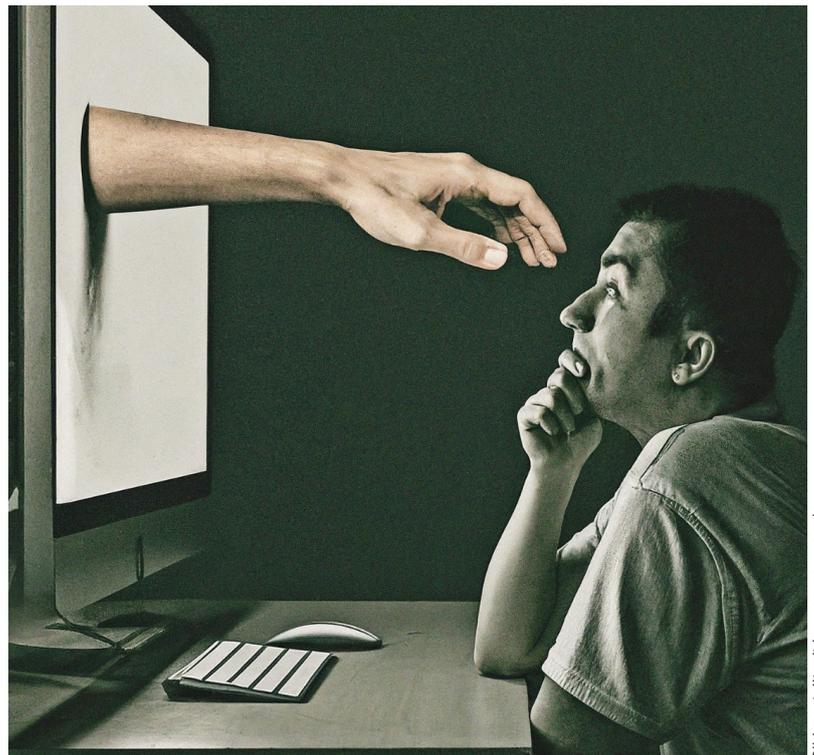


Bild: erstellt mit imagen von google

in der KI widerspiegeln, wenn die Entwicklerinnen und Entwickler Entscheidungen darüber treffen, welche Daten verwendet werden und wie die KI-Modelle konzipiert sind.

Black Box

Viele KI-Systeme sind sogenannte „Black Boxes“, deren Entscheidungsfindungsprozesse nicht transparent oder nachvollziehbar sind. Dies erschwert es, Bias zu erkennen und zu korrigieren, da nicht klar ist, wie die KI zu ihren Schlussfolgerungen kommt.

Verwendete Quellen

Institut für Business Analytics der Universität Ulm: Grundlagen - Bias und Fairness in KI. Bias und Fairness in KI. <https://bias-and-fairness-in-ai-systems.de/grundlagen/>

Graw, M. (31. Oktober 2023): Künstliche Intelligenz und Bias. Mittelstand-Digital Zentrum Zukunftskultur. <https://www.digitalzentrum-zukunftskultur.de/artikel/kuenstliche-intelligenz-und-bias-11821/>

Wood, T. (5. Dezember 2020): Voreingenommenheit in der KI: Wie sich die algorithmische Voreingenommenheit der KI auf die Gesellschaft auswirkt. Fast Data Science. <https://fastdatascience.com/de/vorurteil-in-ki-algorithmus-gesellschaft/>