





Bild: Bruce Rolff / shutterstock

Wie wir KI-Leitlinien im Unternehmen definiert haben

Lisa Stamm – Produktmanagerin bei data42

12.10.2023 @ Remote Brown Bag ARIC

Ethische und gesellschaftliche Herausforderungen von KI und wie wir damit umgehen können

A decorative pattern of small squares in various shades of blue, orange, and grey is scattered across the top half of the slide.

KI-Systeme = sozio-technische Prozesse

Zweig 2019



KI enthält keinen moralischen Kompass



Recruiting



Lerne auf Basis von vorherigen erfolgreichen Einstellungen wer eingestellt werden soll



Amazons bisherige Einstellungen enthielten mehr Männer
Nur falsch positives und kein falsch negatives Feedback
„Women Chess Captain“ wurde negativ bewertet



Gewichtungen in den Daten
Tool wurde eingestellt

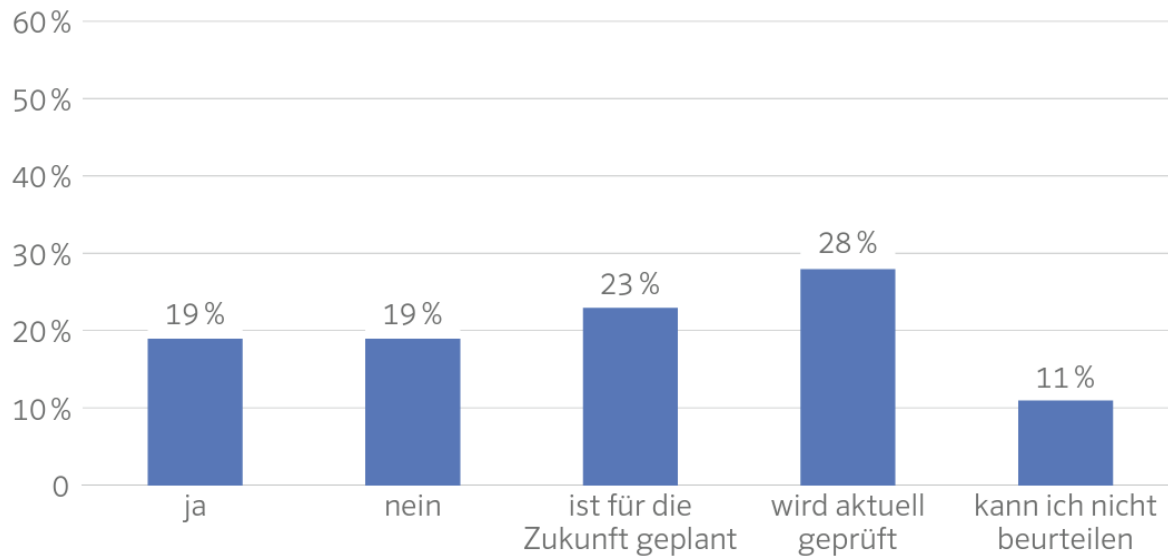
KI-Anwendungen in HR heute ■

- • “Keine automatisierten Entscheidungen”
- • Fokus: “Zusammenfassen der wichtigsten Infos, Erstellen von Inhalten, Onboarding, Mentor Match” ■
- Hinweis: “Enthält keine emotionale Intelligenz”

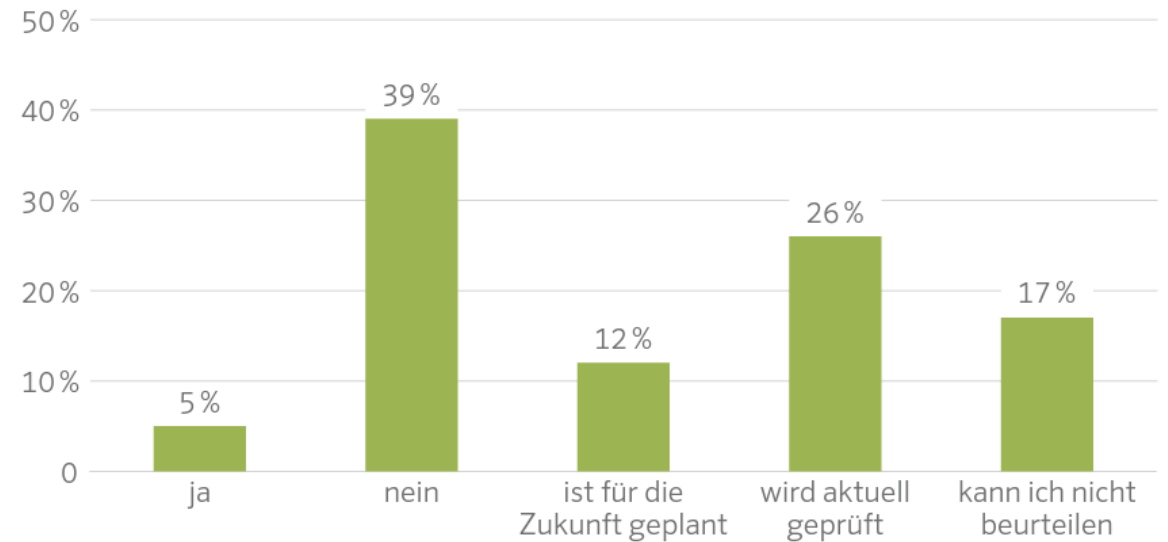
<https://teamflect.com/blog/employee-engagement/ai-in-hr>

Einsatz generativer KI in Deutschland ■

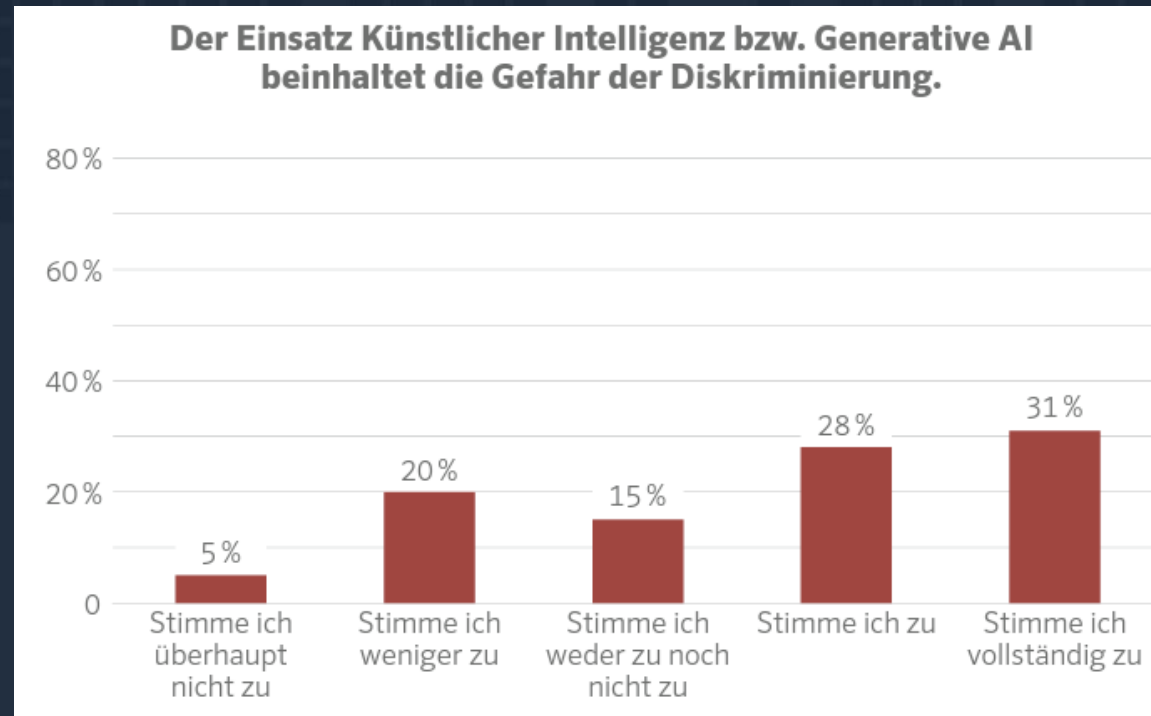
Personalwesen/HR



Bei der Auswahl/Analyse von Bewerbungen



Einsatz generativer KI in Deutschland



59 % sehen die Gefahr von Diskriminierung

Mieter Screening



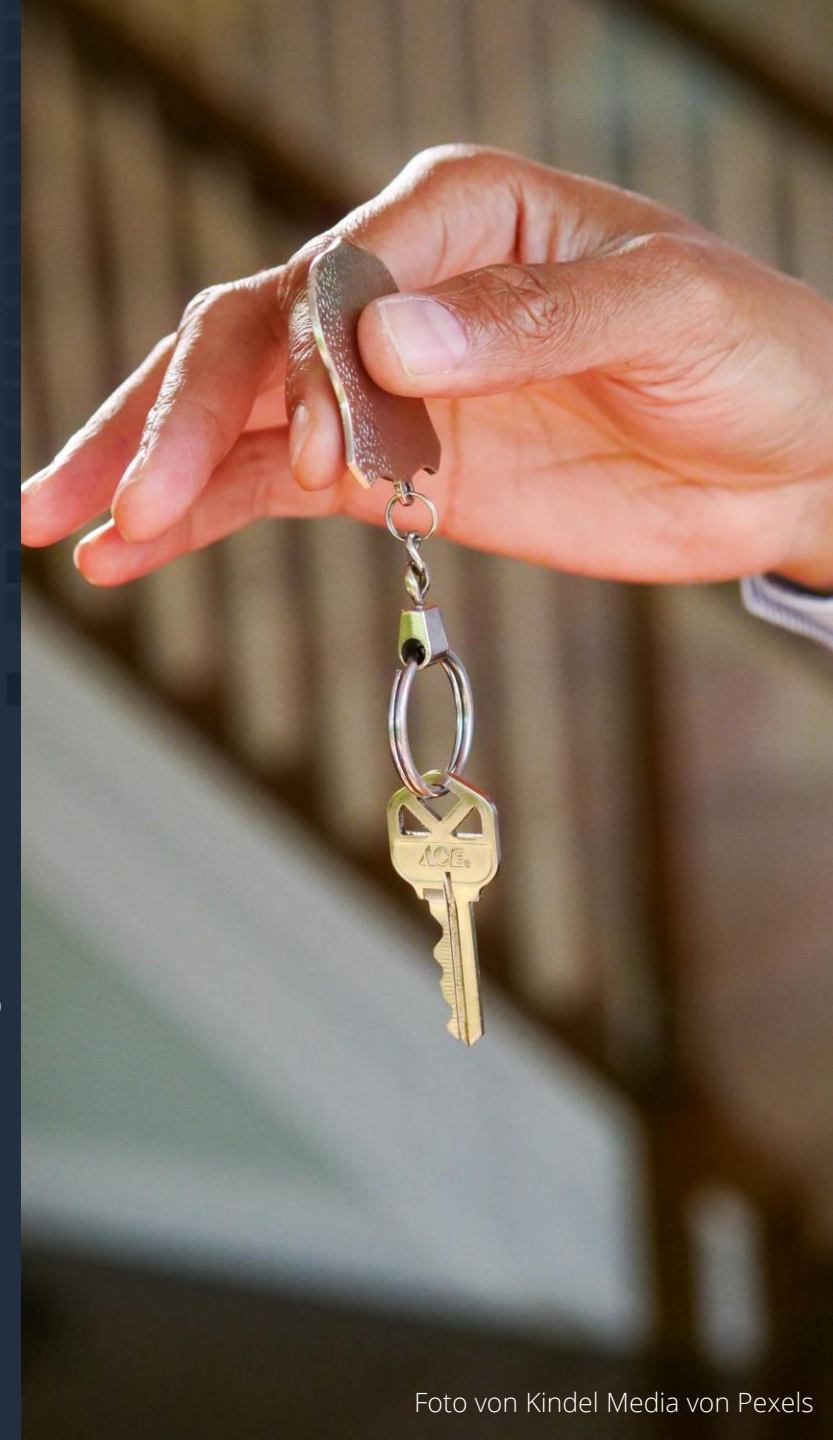
Mieter Screening

- Risikobewertungen enthalten Kriminalhistorie

■ Kritische Information:
Verurteilung/Freispruch fehlt

Unvollständige Daten und lernendes System
führen zu Diskriminierungen

Wonyoung So





Youtube Empfehlungen



Empfehle ein Video was die Nutzerin sieht
Optimiere die Watch Time



Filter Blasen
Radikalisierung



Überarbeitung gegen Radikalisierungsprozesse

Nicht alle KI-Anwendungen sind gesellschaftlich / ethisch problematisch

- Erkennen der Problematik des Anwendungsfalls
 - Starke Auswirkungen auf Menschen
 - Geringe individuelle Auswirkung, aber möglicher gesamtgesellschaftlicher Schaden



Chats zur Entlastung bei Kundenanfragen

> 1000 Anfragen pro Tag
Ca. 80 % wiederholende Fragen

Generative Chatbots liefern Antworten
aus internen & externen Informationen
Mitarbeiter haben mehr Zeit für die
Gäste vor Ort

Beispiel Dialog Shift – Olga Heuser

Generelle Herausforderungen.

- - Unvollständige, verzerrte und falsche Daten
 - Qualitätsmessungen: Accuracy wichtiger als Fairness
 - Unbeabsichtigte, schädliche Optimisierungsnebeneffekte
 - Annahmen ignorieren den Kontext
 - Kreative KI kreiert inkorrekte Informationen

Was können wir tun?

- • Frage: Gibt es ein ethisches / gesellschaftliches Risiko bei dieser Anwendung?
- • Falls ja: Gibt es Maßnahmen, die wir tätigen können, um dieses zu reduzieren?
- • Falls nein: Gibt es alternative Algorithmen/ Werkzeuge, die weniger Risiko bedeuten?

Mögliche Maßnahmen

- • Kontrolle der Daten
- • Gewichtung gegen Gruppen-Biases
- Fokus auf multiple Qualitätsmessungen
- Interdisziplinäre Teams checken mögliche Risiken
- Einbeziehung von vulnerablen Gruppen
- Transparenz von Ziel und Kontext
- Explainable KI nutzen

Beispiel Leitprinzipien

- - Eindeutige Nutzerführung und –verantwortung
 - Transparenz der Algorithmen
 - Datenschutz & -sicherheit



Vielen Dank und frohes Schaffen!

Lisa Stamm von data42

