

Checkliste Risikoabschätzung von KI-Anwendungen am Arbeitsplatz

Künstliche Intelligenz ist ein Werkzeug, das der Verbesserung der Arbeitsqualität dienen soll. Es ist also einerseits festzulegen, wofür KI-Systeme entwickelt werden sollen, aber auch für was sie keinesfalls eingesetzt werden sollen. Dafür sind „rote Haltelinien“ einzuziehen: Es darf keine KI-Anwendungen geben, die Menschen schaden, die gegen Grundrechte verstoßen. Bei einem zu hohen Risiko, wenn ein Schaden nicht ausgeschlossen werden kann, sind KI-Systeme nicht einzusetzen. Beispielsweise sind Profiling-Systeme – also Verfahren, die autonom messen, detektieren, klassifizieren, zählen und prüfen – kritisch zu hinterfragen. Vor der Einführung von KI-Anwendungen am Arbeitsplatz ist daher zu prüfen, ob und welche Risiken für die betroffenen Arbeitnehmer:innen vorliegen. Die vorliegende Checkliste soll Betriebsräte dabei unterstützen, eine Risikoeinschätzung aus Arbeitnehmer:innensicht durchzuführen.

Grundsätze¹, die dabei zu berücksichtigen sind:

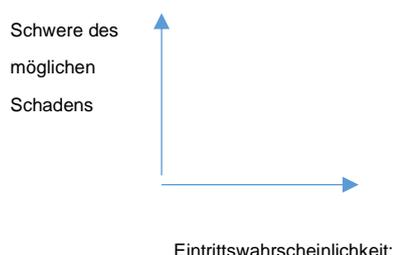
- Transparenz des KI-Systems (Erklärbarkeit und Nachvollziehbarkeit des Ergebnisses)
- Vorrang menschlichen Handelns und menschlicher Aufsicht
- Vielfalt, Nicht-Diskriminierung und Fairness
- Privatsphäre und Datenqualitätsmanagement
- Robustheit von Eingangsdaten und Algorithmen
- Gesellschaftliches und ökologisches Wohlergehen
- Rechenschaftspflicht

Informationen beschaffen und bewerten:

Die Beschreibung der Verarbeitungsvorgänge der geplanten KI-Anwendung ist beim Arbeitgeber einzuholen. Detaillierte Fragen dazu finden sich im GPA-Leitfaden für Betriebsräte zur Einführung von Systemen mit sogenannter „Künstlicher Intelligenz (KI)“

Liegen die nötigen Informationen vor, kann der Betriebsrat die Risiken für die Rechte und Freiheiten der betroffenen AN erkennen, analysieren und bewerten, um die Notwendigkeit und Verhältnismäßigkeit der Anwendung zu prüfen.

Risikomatrix zur Bewertung der Auswirkungen auf die Arbeitnehmer:innen:



¹ Ethik-Leitlinien für eine vertrauenswürdige KI, Hochrangige Expert:innengruppe für KI, EU-Kommission 2019; Europäische Kommission, Weißbuch zur Künstlichen Intelligenz – Ein europäisches Konzept für Exzellenz und Vertrauen; https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf

Mögliche Klassifizierung von KI-Systemen²:

| Kategorie | Schadensrisiko | Charakter/Zweck der Empfehlung, Beispiele | Verhaltens- und Leistungskontrolle |
|-----------|---|--|------------------------------------|
| 1 | Kein Risiko | Unverbindlicher Vorschlag, dessen Befolgen oder Nichtbefolgen keine messbaren Auswirkungen hat. Beispiel: Buchempfehlung | keine |
| 2 | Geringes Risiko | Vorschläge für durchzuführende/nicht durchzuführende Aktionen, Tätigkeiten oder Schulungen als vorbereitende Diskussionsgrundlage zwischen den beteiligten Funktionen. Beispiel: Schulungsvorschläge | keine |
| 3 | Mittleres Risiko (geringe Wahrscheinlichkeit des Eintritts des Schadens, nicht unerhebliche Schadenshöhe möglich) | Vorschläge für durchzuführende/nicht durchzuführende Aktionen, Tätigkeiten oder Schulungen, die einen Einfluss auf die Arbeitsweise/Tätigkeit oder Einstufung der Arbeitnehmer:innen haben. Beispiel: Karriereplanung | möglich |
| 4 | Hohes Risiko (höhere Wahrscheinlichkeit des Eintritts des Schadens, erhebliche Schadenshöhe möglich) | Vorschläge für durchzuführende/nicht durchzuführende Aktionen, Tätigkeiten oder Schulungen, die einen starken Einfluss auf die Arbeitsweise/Tätigkeit oder Einstufung der AN haben, Personalmaßnahmen mit Schadensminimierung für die AN. Beispiel: Automatisierte Auswahl von AN | möglich |
| 5 | Sehr hohes Risiko | Ohne Schadensminimierung für die AN: Personalmaßnahmen, automatisierte Entscheidungen mit unmittelbarer Auswirkung ohne Vorabkontrolle durch Menschen | sicher |

² Orientiert an Checkliste aus IBM-Betriebsvereinbarung, in: Lothar Schröder/Petra Höfers, Praxishandbuch Künstliche Intelligenz, Frankfurt am Main, 2022, S. 400 ff.