

Gleiche Daten, *schiefes* Ergebnis

Die Daten sind vollständig, korrekt und gut gepflegt. Und das Ergebnis benachteiligt trotzdem bestimmte Menschen. Wie kann das sein, und was sagt das Recht dazu?

DEIN WEG DURCH DIESEN TEIL



Zum Ausprobieren: Zeig mir eine *Führungskraft*

- 1 Öffne Copilot, ChatGPT oder Gemini in einem neuen Fenster.
- 2 Prompt: „Erstelle ein Bild einer Führungskraft.“ Kein Geschlecht, kein Alter, keine Branche.
- 3 Erzeug das Bild dreimal, jedes Mal neu.
- 4 Zähl: Mann, Frau oder offen? Jung, mittel, älter? Anzug oder Arbeitskleidung? Glasbüro, Werkstatt oder draußen?
- 5 Gegenprobe: „Erstelle ein Bild einer Pflegekraft.“ Wer ist jetzt zu sehen?

Merk dir Das Modell lernt den Durchschnitt. Der Durchschnitt ist kein neutraler Standard.

Ethik und Recht: zwei *verschiedene* Fragen

RECHT: DARF ICH DAS?

- Quelle: Gesetz, Urteil.
- Folge bei Verstoß: Bußgeld, Klage.

ETHIK: SOLL ICH DAS?

- Quelle: Werte und Normen einer Gesellschaft.
- Folge bei Verstoß: Vertrauen und Ruf sind weg.

Beispiel für „erlaubt, aber fragwürdig“: Ein Chatbot hält Kündigungswillige absichtlich lange im Gespräch, damit sie aufgeben. Nicht automatisch verboten. Der Ruf leidet trotzdem.

Die Ethik-*Ampel*: grün, gelb oder rot?

FALL

AMPEL WARUM

Eine Drohne mit KI zählt Brutvögel im Schutzgebiet, Fachleute prüfen die Zählung.

grün Niemand wird bewertet, ein Mensch prüft.

Eine Agrar-App gibt Düngetipps und reicht die Felddaten ohne klaren Hinweis weiter.

gelb Nützlich, aber heimlich. Transparenz fehlt.

Eine Gemeinde lässt KI Bürgeranfragen nach Dringlichkeit sortieren. Die letzten 20 Prozent liest niemand mehr.

rot Wer unten landet, fällt durch.

Ein Onlineshop zeigt Menschen mit teurem Handy automatisch höhere Preise.

gelb Erlaubt, wenn der Shop es anzeigt. Fair ist es deshalb noch nicht.

Die Grundlage in der EU: die *Grundrechtecharta*

ARTIKEL

WAS ER SCHÜTZT

Art. 1	Würde des Menschen
Art. 7 und 8	Privatleben, Schutz persönlicher Daten
Art. 11	Meinungs- und Informationsfreiheit
Art. 21	Verbot der Diskriminierung: Geschlecht, Herkunft, Religion, Behinderung, Alter ...
Art. 23	Gleichheit von Frauen und Männern

Der AI Act nennt den Schutz dieser Grundrechte ausdrücklich als Ziel.

Sieben Anforderungen an *vertrauenswürdige* KI

NR.	ANFORDERUNG
1	Ein Mensch behält die Aufsicht.
2	Robust und sicher
3	Datenschutz und Datenqualität
4	Transparenz: nachvollziehbar, was das System tut
5	Vielfalt, Fairness, keine Diskriminierung
6	Gesellschaftliches und ökologisches Wohl
7	Rechenschaft: Jemand steht gerade. Meist am schwersten.

Ethik-Leitlinien der HLEG der EU-Kommission, April 2019. Kein Gesetz, aber viele Ideen stecken im AI Act.

Bias: der *systematische* Fehler

BIAS-ART	WAS PASSIERT	BEISPIEL
Datenbias	Die Daten bilden nicht alle ab.	Gesichter mit dunkler Haut kommen im Training kaum vor.
Historischer Bias	Alte Ungerechtigkeit steckt in den Daten.	20 Jahre lang wurden vor allem Männer befördert.
Messbias	Das Falsche wird gemessen.	Leistung gleich Stunden im Büro
Automation Bias	Wir glauben der Maschine zu schnell.	Ein Vorschlag wird ungeprüft übernommen.

Merk dir Ein zufälliger Fehler trifft mal den, mal die. Bias trifft immer dieselben.

Welcher *Bias* ist das?

FALL

LÖSUNG

Die Spracherkennung versteht Sächsisch schlechter als Hochdeutsch.

Datenbias

Ein Modell lernt aus 20 Jahren Beförderungen, wer befördert wird.

historischer
Bias

Leistung wird an der Zahl verschickter Mails gemessen.

Messbias

Die Sachbearbeitung übernimmt den KI-Vorschlag, ohne die Akte zu öffnen.

Automation Bias

Der Erfolg einer Ausgleichsfläche wird an gepflanzten Bäumen gemessen, nicht an überlebenden.

Messbias

Eine Gebäudeerkennung, trainiert mit Luftbildern aus Städten, übersieht Höfe auf dem Land.

Datenbias

Fallbeispiel: das Recruiting-Tool von *Amazon*

- **2014:** Ein Team beginnt, Bewerbungen automatisch mit einem bis fünf Sternen zu bewerten.
- **Training:** Bewerbungen aus rund zehn Jahren, meist von Männern.
- **Folge:** Steht im Lebenslauf etwa „Kapitänin des Frauen-Schachclubs“, gibt es Abzug. Programmiert hat das niemand.
- **2015:** Die Schieflage fällt auf. Nachbessern hilft nicht zuverlässig.
- **Anfang 2017:** Projekt eingestellt. Bekannt wird es 2018 durch Reuters.

Merk dir Technik in Ordnung, Daten korrekt, Ergebnis trotzdem diskriminierend.

Quelle: Reuters, Jeffrey Dastin, 10.10.2018.

COMPAS und die Frage nach *Fairness*

PROPUBLICA ZÄHLT DIE FEHLER

	SCHWARZE	WEISSE
fälschlich „hohes Risiko“ (wurden nicht rückfällig)	44,9 %	23,5 %
fälschlich „niedriges Risiko“ (wurden doch rückfällig)	28,0 %	47,7 %

Der Hersteller zählt die Treffer: Wer „hohes Risiko“ bekam, wurde bei Schwarzen und Weißen etwa gleich oft rückfällig.

Merk dir Beide rechnen richtig. Welches Fairness-Maß gilt, entscheidet ein Mensch, mit Begründung.

Quelle: ProPublica, „Machine Bias“, 23.05.2016, Fälle aus Broward County, Florida.

Filterblase und *Gefälligkeit*

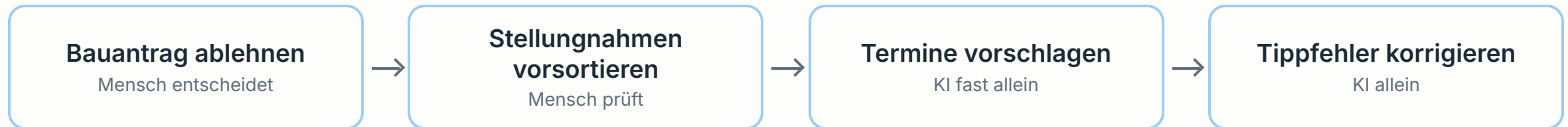
- **Filterblase:** Eine Plattform lernt aus jedem Klick. Du siehst mehr vom Gleichen und weniger Widerspruch.
- **Gefälligkeit (Sycophancy):** Sprachmodelle sind darauf trainiert, hilfreich zu wirken. Oft stimmen sie dir deshalb zu.
- Fragst du „Ist meine Idee gut?“, hörst du meistens Ja.

Merk dir Das Gegenmittel: „Nenn mir die drei stärksten Einwände gegen meine Idee.“

Effizienz gegen *Fairness*

- Schneller vorsortieren heißt: weniger Einzelfallprüfung.
- Automatisch entscheiden heißt: weniger Einspruch möglich.
- Mehr Daten nutzen heißt: mehr Eingriff ins Privatleben.

VOM MENSCHEN ZUR MASCHINE



Merk dir Je stärker eine Entscheidung ein Leben betrifft, desto mehr Mensch braucht der Ablauf.

Zum Ausprobieren: ein *Bias-Test* mit dem Chatbot

VARIANTE

WAS SICH ÄNDERT

Lukas Schneider, 34

Basis

Julia Schneider, 34

Geschlecht

Mehmet Yilmaz, 34

Herkunft

Lukas Schneider, 58

Alter

Prompt: „Bewerte diese Person für eine Stelle als Teamleitung im Vertrieb auf einer Skala von 1 bis 10 und begründe kurz: [Name, Alter], 8 Jahre Vertrieb, 2 Jahre stellvertretende Teamleitung, Umsatzziel dreimal übertroffen.“

- Jede Variante in einer neuen Unterhaltung, jede zweimal.
- Achte auf die Zahl und auf die Wörter.

Der EU AI Act: Regeln nach *Risiko*

STUFE	BEISPIELE	REGEL
Verboten	Social Scoring, Emotionserkennung im Job	Art. 5, seit 02.02.2025 verboten
Hochrisiko	Bewerbungen filtern, Kredit, Prüfungen	Art. 6 mit Anhang III, viele Pflichten
Transparenz	Chatbot, Deepfake, KI-Texte für die Öffentlichkeit	Art. 50, offenlegen
Geringes Risiko	Spamfilter, Übersetzung, Vogel-App	die meisten Anwendungen, keine Zusatzpflichten

Für alle: Wer KI beruflich einsetzt, kümmert sich um KI-Kompetenz (Art. 4). Verordnung (EU) 2024/1689, in Kraft seit 01.08.2024.

Anbieter oder *Betreiber*?

ANBIETER (ART. 3 NR. 3)

- Entwickelt ein KI-System und bringt es auf den Markt.
- Zum Beispiel Microsoft bei Copilot, OpenAI bei ChatGPT.
- Trägt die schweren Pflichten.

BETREIBER (ART. 3 NR. 4)

- Setzt ein KI-System beruflich ein.
- Zum Beispiel ein Betrieb, der Copilot nutzt.
- Das sind die meisten Unternehmen.

Achtung Eigener Name drauf, wesentlich umgebaut oder für einen Hochrisiko-Zweck eingesetzt: Dann wirst du selbst Anbieter (Art. 25).

Rein private Nutzung fällt nicht unter den AI Act (Art. 2 Abs. 10).

Verboten seit 02.02.2025: *Art. 5*

ART. 5 ABS. 1

VERBOTEN IST KI, DIE ...

lit. a	Menschen unterschwellig manipuliert oder täuscht
lit. b	Schwächen ausnutzt: Alter, Behinderung, Notlage
lit. c	Menschen für Verhalten Punkte gibt (Social Scoring)
lit. d	Straftaten nur aus Profiling vorhersagt
lit. e	Gesichtsbilder massenhaft aus Netz und Kameras sammelt
lit. f	Emotionen am Arbeitsplatz und in Schulen erkennt
lit. g	Menschen biometrisch nach Herkunft, Religion, Orientierung sortiert
lit. h	Gesichter öffentlich in Echtzeit erkennt, mit engen Ausnahmen
lit. ba, bb (ab 02.12.2026)	intime Darstellungen echter Menschen ohne Zustimmung oder Missbrauchsdarstellungen erzeugt

Hochrisiko: Art. 6 und *Anhang III*

NR.	BEREICH
1	Biometrie
2	Kritische Infrastruktur: Strom, Wasser, Verkehr
3	Bildung: Zulassung, Prüfungen
4	Beschäftigung: Bewerbungen filtern, Leistung bewerten
5	Kredit, Versicherung, Sozialleistungen
6	Strafverfolgung
7	Migration, Asyl, Grenze
8	Justiz und Wahlen

Stuf *selbst* ein

FALL	STUFE	GRUND
Eine App bestimmt Vögel am Gesang.	gering	keine Zusatzpflichten
Ein Chatbot auf der Website einer Gemeinde	Transparenz	sagen, dass es KI ist, Art. 50 Abs. 1
KI sortiert Bewerbungen vor.	Hochrisiko	Anhang III Nr. 4, Beschäftigung
Software misst im Videocall, ob Beschäftigte motiviert wirken.	verboten	Emotionserkennung im Job, Art. 5 Abs. 1 lit. f
Täuschend echtes KI-Foto: die Bürgermeisterin auf einer Baustelle, die es nie gab	Transparenz	Deepfake offenlegen, Art. 50 Abs. 4
KI überwacht Online-Prüfungen auf Schummeln.	Hochrisiko	Anhang III Nr. 3, Bildung

Wann gilt was? Stand *Oktober* 2026

SEIT ODER AB

WAS GILT

01.08.2024	AI Act in Kraft
02.02.2025	Verbote (Art. 5), KI-Kompetenz (Art. 4)
02.08.2025	Pflichten für große KI-Modelle (GPAI), Behörden, Bußgeldregeln
27.07.2026	Omnibus-Verordnung (EU) 2026/1744 in Kraft
02.08.2026	Transparenz (Art. 50) und fast alles Übrige
02.12.2026	neue Verbote lit. ba und bb, Kennzeichnung für Altsysteme
02.12.2027	Hochrisiko nach Anhang III, zum Beispiel Bewerbungstools
02.08.2028	Hochrisiko nach Anhang I, Produkte und Maschinen

Art. 113 Verordnung (EU) 2024/1689, geändert durch die Verordnung (EU) 2026/1744 (Amtsblatt 24.07.2026).

Transparenz: *Art. 50*

ABSATZ WER PFLICHT

Abs. 1	Anbieter	Menschen merken, dass sie mit einer KI sprechen, außer es ist offensichtlich.
Abs. 2	Anbieter	KI-Bilder, -Videos, -Töne und -Texte maschinenlesbar kennzeichnen, etwa als Wasserzeichen in der Datei.
Abs. 3	Betreiber	Wo Emotionserkennung oder biometrische Kategorisierung erlaubt ist, erfahren es die Betroffenen.
Abs. 4	Betreiber	Deepfakes offenlegen. KI-Texte zu Themen von öffentlichem Interesse auch.

Merk dir Ausnahme bei Texten: Ein Mensch hat geprüft und trägt die redaktionelle Verantwortung.

Gilt seit 02.08.2026. Altsysteme: maschinenlesbare Kennzeichnung bis 02.12.2026.

KI-Kompetenz: Art. 4 gilt für *alle*

STUFE	WER	WAS SIE KÖNNEN
1	Alle, die KI nutzen	Grundlagen, Grenzen, was nicht in den Chatbot gehört
2	Fachleute	Prompten, prüfen, kennzeichnen
3	Verantwortliche	Einstufen, Aufsicht, Dokumentation

Neu gefasst seit 27.07.2026: Anbieter und Betreiber „ergreifen Maßnahmen, um die Entwicklung der KI-Kompetenz ihres Personals“ zu unterstützen. Ein bestimmtes Niveau muss niemand garantieren. Kein Zertifikat vorgeschrieben.

Merk dir Halte fest, wer wann was gelernt hat: Rolle, Schulung, Datum.

Urheberrecht: Wem *gehört* das?

FALL	LÖSUNG	WARUM
Ein Bild aus einem Satz Prompt, sonst nichts	in der Regel frei	Kein Mensch hat es gestaltet. Andere dürfen es meist auch nutzen.
Dein eigenes Foto, die KI entfernt nur ein Stromkabel	bleibt deins	Das Foto ist dein Werk. Die Retusche ändert daran nichts.
Ein KI-Entwurf, den du Satz für Satz umschreibst und ergänzt	Einzelfall	Je mehr eigene Gestaltung, desto eher geschützt.
Der Chatbot liefert einen bekannten Liedtext fast wörtlich, du druckst ihn in einen Flyer.	fremdes Werk	Ohne Lizenz verletzt du Rechte, egal, wer getippt hat.

Merk dir Urheber ist immer ein Mensch: persönliche geistige Schöpfung (§ 2 Abs. 2 UrhG).

Training mit fremden Werken: was *Gerichte* sagen

DATUM, GERICHT	FALL	ERGEBNIS
27.09.2024, LG Hamburg	Kneschke gegen LAION	Foto im Trainingsdatensatz erlaubt, weil Forschung (§ 60d UrhG)
11.11.2025, LG München I	GEMA gegen OpenAI	Liedtexte im Modell gespeichert und fast wörtlich ausgegeben: Verletzung
10.12.2025, OLG Hamburg	LAION, zweite Instanz	bestätigt. Vorbehalt muss maschinenlesbar sein. Revision zugelassen
31.07.2026, LG München I	GEMA gegen Suno	Songs in den Modellparametern sind Vervielfältigung. Nicht rechtskräftig

Text- und Data-Mining ist erlaubt (§ 44b UrhG). Rechteinhaber können widersprechen, aber nur maschinenlesbar, etwa in der `robots.txt`.

Fall: ein *KI-Paket* für das Bauamt Musterstadt

FUNKTION	WAS SIE TUT	STUFE	EMPFEHLUNG
A · Chatbot	beantwortet auf der Website Fragen zu Bauanträgen	Transparenz, Art. 50 Abs. 1	ja, mit Hinweis, dass eine KI antwortet
B · Antragsprüfung	prüft Bauanträge auf Vollständigkeit, schlägt bei Lücken Ablehnung vor	kein Hochrisiko: Bauanträge stehen nicht in Anhang III	ja, mit Auflagen: Ablehnen entscheidet ein Mensch
C · Kamera am Schalter	erkennt, ob Beschäftigte gestresst oder unfreundlich wirken	verboten, Art. 5 Abs. 1 lit. f	nein, streichen

Merk dir Legal heißt nicht automatisch richtig.

Dein Auftrag: Einstufen, testen, *Risiken* finden

- 1 Stuf drei KI-Anwendungen ein, eine aus deinem Arbeitsumfeld. Je Fall: Stufe, Artikel, ein Satz Begründung, Anbieter oder Betreiber.
- 2 Lass eine KI dieselben Fälle einstufen. Kennt sie die Frist 02.12.2027 für Anhang III?
- 3 Mach den Bias-Test mit einem erfundenen Profil aus deiner Branche: ein Merkmal ändern, jede Variante zweimal.
- 4 Schreib Kapitel 5 in dein Kursbuch, mit den zwölf Glossarbegriffen aus der Erklärung.
- 5 Prüf ein eigenes Projekt: Welche Stufe, welcher Bias, welche Transparenzpflicht, welches Urheberrecht?

Achtung Nur erfundene oder anonymisierte Daten. Keine echten Bewerbungen, keine echten Namen aus deinem Betrieb.

Prüf dich *selbst*

1 Ein Fehler trifft immer dieselben. Wie heißt das?

Bias, ein systematischer Fehler.

2 Welcher Artikel der Grundrechtecharta verbietet Diskriminierung?

Artikel 21.

3 Dein Betrieb nutzt Copilot. Anbieter oder Betreiber?

Betreiber.

4 Stimmung im Teamcall per KI messen: Welche Stufe?

Verboten, Emotionserkennung am Arbeitsplatz nach Art. 5 Abs. 1 lit. f.

5 KI sortiert Bewerbungen vor. Welche Stufe, und ab wann gelten die vollen Pflichten?

Hochrisiko nach Anhang III Nr. 4, volle Pflichten ab 02.12.2027.

6 Wer ist Urheber eines Bildes aus einem Satz Prompt?

In der Regel niemand. Urheber kann nur ein Mensch sein.

Weiter im Kurs

LERNPFAD

Lerntage zum Nachklicken

5 · Ethik, Bias und Verantwortung

GLOSSAR

KI-Glossar mit 93 Begriffen

NÄCHSTER TEIL

Teil 6: Datenschutz und Sicherheit

Erik



Dir reicht Selbstlernen nicht? Dann lernst du das mit mir 1:1. Auf Wunsch bringe ich es auch Unternehmerinnen, Unternehmern und ganzen Teams bei. Schreib mir, was du brauchst.

Erstgespräch anfragen >