



Bundesnetzagentur



Netzwerktreffen | KI & Ethik

25. März 2026
9:00 – 10:30 Uhr

Agenda:

1. **Begrüßung und Einführung**
2. **Was ist ethische KI – und wie setzen Unternehmen sie um?**
Lajla Fetic, appliedAI Institute for Europe
3. **Algorithmische Diskriminierung in der Unternehmenspraxis: Risiken und Handlungsstrategien**
Lejla Eva Podgoršek, AlgorithmWatch
4. **Vertrauen und Verantwortung in Zeiten von Agentic AI – Erkenntnisse und Praxisbezüge**
Janek Kuberzig, Bundesverband Digitale Wirtschaft (BVDW) e.V.

Vielen Dank für Ihre Aufmerksamkeit!

Bei Fragen oder weiterem Interesse melden Sie sich gerne unter:
Digitalisierung_KMU@BNetzA.DE

Hier finden Sie außerdem den Link zur **Anlaufstellendatenbank**:

<https://www.bundesnetzagentur.de/DE/Fachthemen/Digitales/Mittelstand/Anlaufstellen/start.html>

Was ist „ethische KI“ und wie gestalten wir sie?

Lajla Fetic,
Head of Public Interest & Policy

A joint
initiative

UNTER
NEHMER
TUM

 IPAI

**Nicht die Technologie
entscheidet über ihren
gesellschaftlichen Wert,
sondern unsere Gestaltung.**

Wieso ist Ethik bei der Gestaltung und dem Einsatz von KI (noch) wichtig - trotz KI-Verordnung?

Ethik verlangt uns - unabhängig von Prozessen und Strukturen Reflexionsfähigkeiten ab.



KI ist nicht neutral. Aber auch nicht böse und nicht gut.

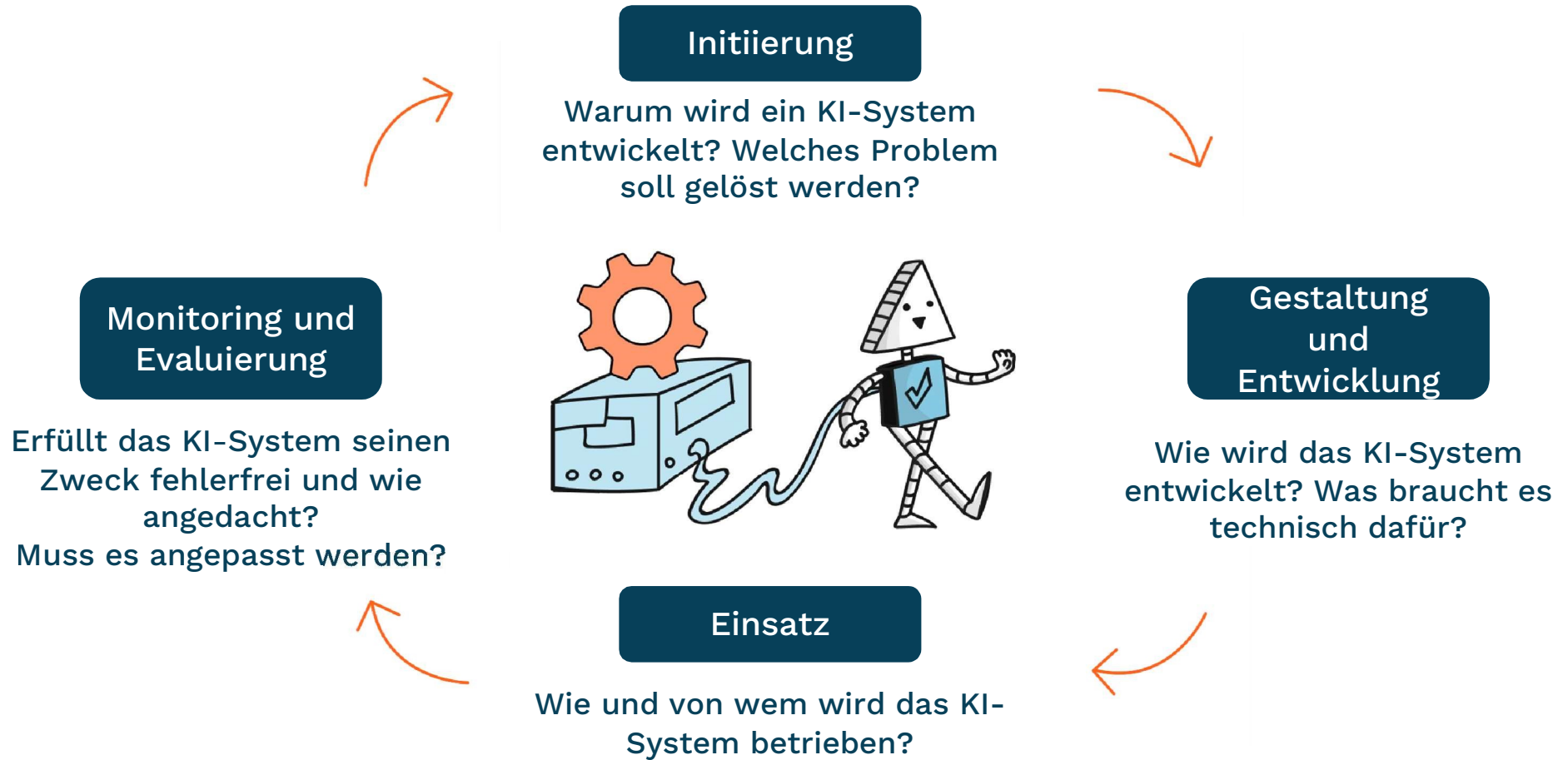
Ihr Einsatz hält uns einen Spiegel unserer Werte vor.



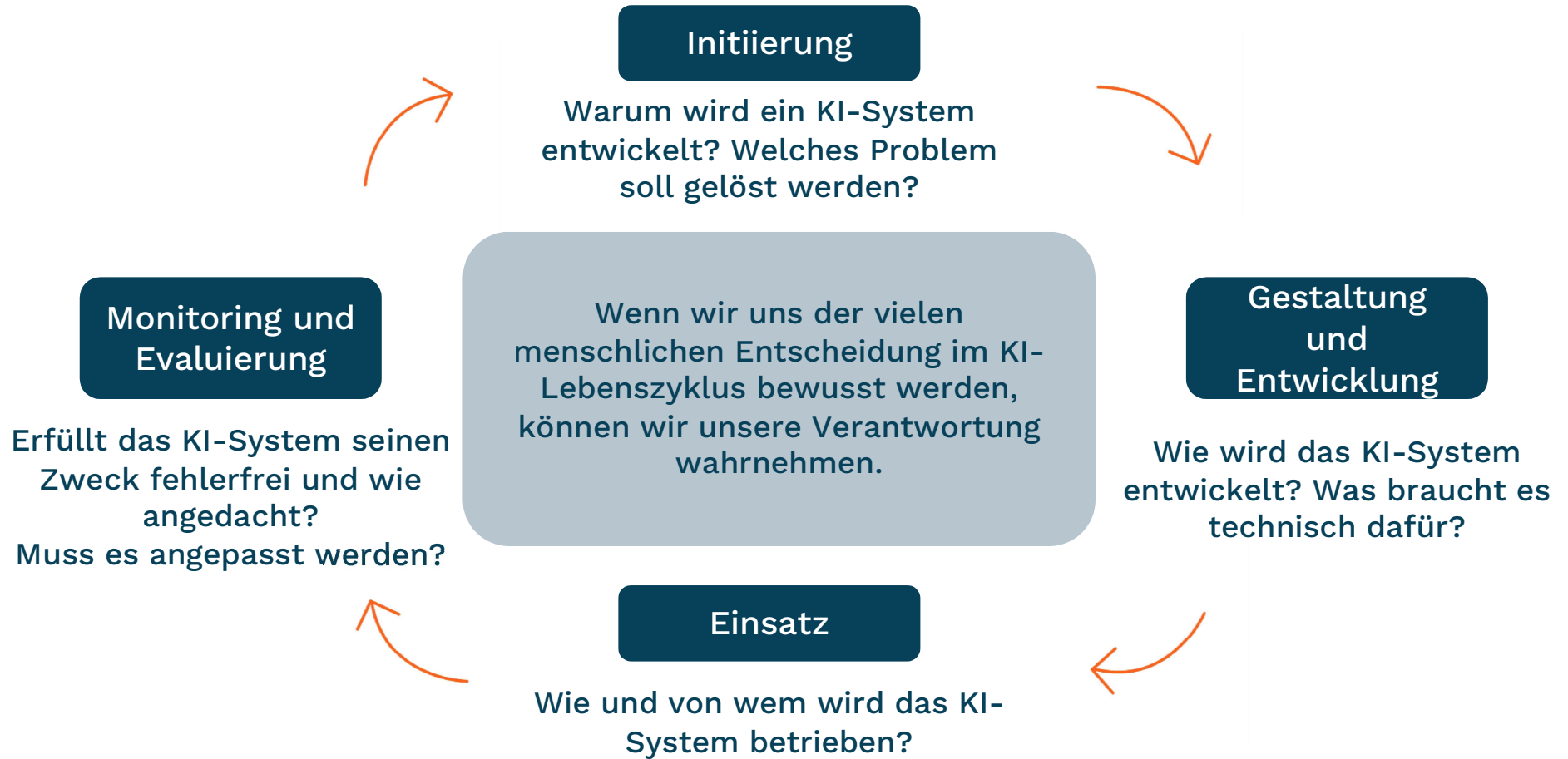
KI ist mehr als die technischen Komponenten. Viele menschliche Entscheidungen prägen die Entwicklung und den Einsatz.



Was ist der KI-Lebenszyklus?



Was ist der KI-Lebenszyklus?



Entlang des KI-
Lebenszyklus können
wir auf Basis ethischer
Prinzipien fragen, was
richtige,
verantwortungsvolle
Entscheidungen sind.

Accountability



Privacy



Justice



Beispiel: Der Grundsatz der Selbstbestimmung und die Zielsetzung für den Einsatz von KI-Systemen

Welchen Eingriff in die Autonomie des Menschen wollen wir zulassen? Welchen beschränken?



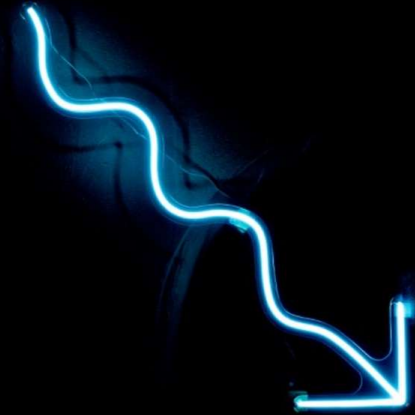
Beispiel: Fairness bei der technischen Entwicklung von KI-Systemen

Welche Vielfalt der Spracheingabe - und ausgabemöglichkeiten bei Chatbots möchten wir darstellen? Welche limitieren?



Beispiel:
Nachvollziehbarkeit im
Betrieb von KI-
Systemen

Wie nachvollziehbar
müssen die
Filterkriterien für
Lebenslauf-
Analysesoftware sein?



Beispiel: Robustheit bei
der Evaluation von KI-
Systemen

Welche Fehlerrate
akzeptieren wir bei der
KI-gestützten
Auswertung von
Schutzkleidung?



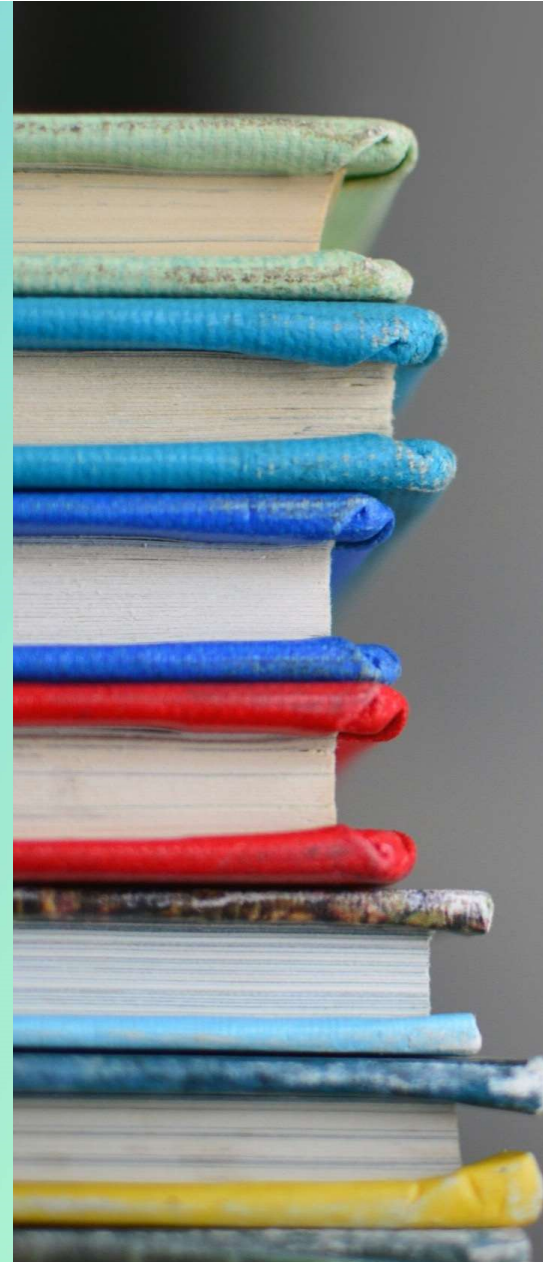
Drei Schritte, wie wir KI-Ethik in unternehmerische Praxis verankern:

1. Verantwortung anerkennen, entlang des gesamten KI-Lebenszyklus
2. Ethische Prinzipien als Leitplanken etablieren
3. Anwendung der Prinzipien operationalisieren durch Recht oder neue Diskursräume



Wieso braucht es
ethische Leitplanken bei
der Gestaltung und dem
Einsatz von KI?

Sie schaffen die
Grundlage für gut
begründetes Vertrauen.



**Nicht die Technologie
entscheidet über ihren
gesellschaftlichen Wert,
sondern unsere Gestaltung.**

Bleiben wir in Kontakt!



Lajla Fetic

Head of AI Public Interest & Policy

l.fetic@appliedai-institute.de

www.appliedai-institute.de | info@appliedai-institute.de





Algorithmische Diskriminierung in der Unternehmenspraxis: Risiken und Handlungsstrategien

LEJLA EVA PODGORŠEK | NETZWERKTREFFEN BUNDESNETZAGENTUR | 25. MÄRZ 2026

algorithmwatch.org



/ AlgorithmWatch





/ AlgorithmWatch

Organisation: Gemeinnützige Nichtregierungs-Organisation (CH & DE)

Fokus: Auswirkungen von Algorithmen und KI auf Menschen und Gesellschaft

Mission: Algorithmen stärken Grundrechte, Demokratie, Gerechtigkeit & Nachhaltigkeit statt Diskriminierung, Machtungleichgewicht und demokratischädliche Debattenkulturen zu befördern

Ansatz: Evidenzbasierte Policy- und Advocacy-Arbeit



/ AlgorithmWatch



Wir streben eine Welt an, in der Technologie im Allgemeinen und algorithmische Systeme im Besonderen den Menschen zugute kommen. Die Systeme sollen Gesellschaften gerechter, demokratischer, inklusiver und nachhaltiger machen – sei es hinsichtlich zugeschriebener Herkunft und Gender, Rassifizierung, sexueller Orientierung, Alter, Klasse und Wohlstand oder Ressourcenverbrauch.



/ Übersicht:

1. Was ist algorithmische Diskriminierung?
2. Ursachen und Beispiele
3. Algorithmische Diskriminierung in der Unternehmenspraxis
4. Was können wir gegen algorithmische Diskriminierung tun?
5. Weiterführende Ressourcen/ Denkanstöße



/ Was ist algorithmische Diskriminierung?



/ Wo entsteht algorithmische Diskriminierung?

Bei der Nutzung **algorithmischer Systeme**, die eingesetzt werden (können),

- für **Entscheidungen** über Menschen
- um Inhalte zu generieren, die **Entscheidungen von/ über Menschen** beeinflussen.

D.h.: Bei der **Automatisierung von Entscheidungen** mit Software (z.B. Künstliche Intelligenz, Plattform-Algorithmen etc.)...

...können algorithmische Systeme Entscheidungen treffen, die Menschen diskriminieren.



/ Definition?

- Kein feststehender Begriff: Umschreibt verschiedene Arten **ungerechtfertigter Benachteiligung**, die im **Zusammenhang mit automatisierten Systemen** auftritt
- **Mensch** spielt dabei eine **zentrale Rolle**
- **Folgen: Verstärkung bestehender Ungleichheiten**

Algorithmische Diskriminierung beschreibt eine **neue Dimension von Ungleichheit**, die oft **schwer sichtbar** und daher **für Betroffene schwierig zu erkennen** ist.



/ Ursachen...



Wie entsteht algorithmische Diskriminierung?

Durch Menschen = Technik ist nie neutral

Automatisierte Entscheidungssysteme sind nicht neutral oder objektiv:

- Daten
- Ziele
- (mathematische) Modelle
- Anwendung

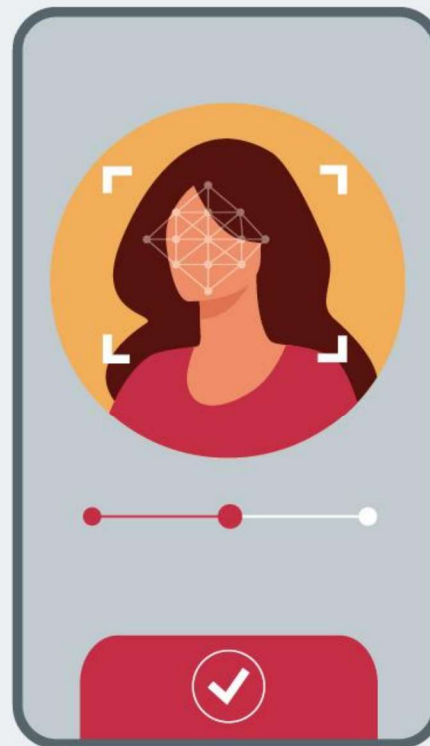


...reproduzieren Vorurteile und Diskriminierungsmuster, die in der Gesellschaft existieren.

DATEN

Sind die Daten repräsentativ für die Gruppen, auf die das ADM-System einen Effekt haben wird?

Wer wählt die Daten aus und warum wählt die Person diese Daten (z. B. Qualität der Daten, Verfügbarkeit oder Kosten)?



Sind die Daten aktuell und ausreichend? Wer prüft das?

Welche Fragen können mit den Daten beantwortet werden?

Wer hat die Daten annotiert und nach welchen Kriterien?



ZIEL

Welcher Zweck soll durch den Einsatz eines ADM-Systems erfüllt werden?

Welche (kritischen) Fragen werden gestellt?
Welche Alternativen in Betracht gezogen?


Welche Annahmen über ADM-Systeme stecken hinter der Einschätzung, dass dies die beste Lösung für das identifizierte Problem ist?

Wer definiert das Ziel des Einsatzes, das Problem und die Lösung? Welche Annahmen hat diese Person?

Welche Perspektiven auf die Gesellschaft hat die Entscheidungsperson?



MODELL



Auf welchen Annahmen über das Problem, die Realität oder das Ziel basiert das Modell?

Wer übersetzt das Problem in ein Modell und wie wird es übersetzt?

Welchen Ausschnitt der Realität bildet das Modell ab? Ist die Vereinfachung dem Problem angemessen?

Ist das Modell geprüft worden? Nach welchen Qualitätsstandards? Unternehmensintern oder von unabhängiger Stelle?

ANWENDUNG

Welche Kompetenzen und welches Wissen haben die Anwendenden?

Haben Anwendende die Möglichkeit oder Freiheit, sich anders zu entscheiden als vom System empfohlen?

Sind Anwendende geschult, das System anzuwenden? Kennen sie die Möglichkeiten und Grenzen?

Wer wendet das ADM-System an und wofür?

Falls formal möglich: Ist dies im Arbeitsalltag realistisch, z. B. angesichts knapper Zeit- oder Personalressourcen? Welche Konsequenzen ergeben sich bei einem Widerspruch für Anwendende?





/ ...und Beispiele algorithmischer Diskriminierung



Online-Handel



Finanzsektor



Gesundheitssektor



Versicherungssektor



Arbeitsmarkt



Bildungssektor



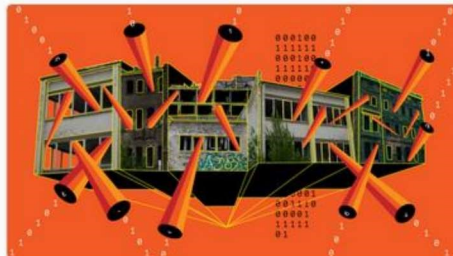
Wohnungssektor



SERIE MIT FALLBEISPIELEN

/ Diskriminierung 2.0: Wie Algorithmen Menschen diskriminieren

Anhand von konkreten Fallbeispielen beleuchten wir unterschiedlichste Fälle von algorithmischer Diskriminierung.



KURZ ERKLÄRT

Wie Algorithmen diskriminieren – und was wir dagegen tun sollten



8. MÄRZ – INTERNATIONALER FRAUENTAG

Wie Algorithmen Frauen diskriminieren



21. MÄRZ – INTERNATIONALER TAG GEGEN RASSISMUS

Wie Rassismus in Algorithmen weiterlebt



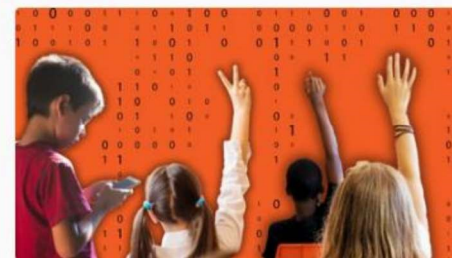
JUNI – PRIDE MONTH

Algorithmen sind binär, Menschen nicht



17. OKTOBER – TAG FÜR DIE BESEITIGUNG DER ARMUT

Wie Algorithmen das Leben von Menschen in Armut erschweren



20. NOVEMBER – TAG DER KINDERRECHTE

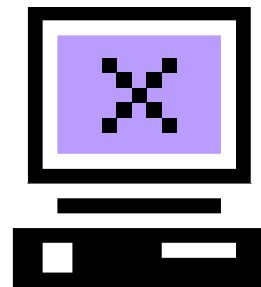
So kann Künstliche Intelligenz Kinder benachteiligen



DE



/ AlgorithmWatch sammelt Fälle von algorithmischer Diskriminierung



**ALGORITHMISCHE
DISKRIMINIERUNG
MELDEN**



/ Algorithmische Diskriminierung in der Unternehmenspraxis

Teams kann demnächst "petzen"

Digitale Überwachung etabliert sich schleichend im Arbeitsalltag

09.11.2025, 17:22 Uhr



Von Diana Dittmer



Mitarbeiterüberwachung führt mitunter zu absurden Gegenmaßnahmen: "Mouse Jiggler" täuschen Mausebewegungen vor. Anwesenheit lässt sich auch simulieren, indem man sich mit Fake-WLANs verbindet. Diese Tricks sind demnächst nicht mehr möglich. (Foto: picture alliance/dpa)



Folgen auf:

In wenigen Wochen führt Microsoft eine neue Teams-Funktion ein, die erkennen kann, ob sich Beschäftigte tatsächlich im Büro aufhalten. Kritiker warnen: Was als Komfortfunktion gedacht ist, öffnet die Tür für neue Formen der Überwachung.



ADVOCACY



Facial Recognition Software Struggles To Detect Dark Skin — Here's Why & How



By Xavier Harding | Aug. 7, 2023

deVolkskrant



The anti-cheating software did not recognize her as a human because she is Black, but she found no understanding at the VU.



Master's student Robin Pocornie (left) and Hans de Zwart from the Racism and Technology Center. Guus Dubbelman / de Volkskrant

Fleur.Damen July 15, 2022, 05:00

PROJEKT «FINDHR»

/ Just Hiring! So vermeiden wir Diskriminierung durch Algorithmen in Job-Bewerbungsverfahren

KI, die Job-Bewerbungen vorsortiert? Im Rahmen des Horizon-Europe-Projekts «FINDHR» haben wir Algorithmen, Methoden und Trainings entwickelt, um Diskriminierung zu reduzieren, wenn KI in der Personalauswahl zum Einsatz kommt. Hier sind unsere Ergebnisse.



DE





A TOOLKIT FOR HR PROFESSIONALS

HOW TO COUNTER DISCRIMINATION IN AI-ASSISTED RECRUITING WITHIN YOUR ORGANISATION.

The illustration shows three diverse people in professional attire. The woman in the center is holding a large, glowing circular graphic. Below them are several blue server racks. The background is dark with various icons representing AI and recruitment.

EN



A TOOLKIT FOR SOFTWARE DEVELOPERS

HOW TO COUNTER DISCRIMINATION IN AI-POWERED RECRUITING SOFTWARE.

The illustration shows four diverse people sitting around a table, working on laptops. A large, glowing blue face is superimposed over the scene, with various icons representing AI and recruitment. The background is dark with various icons representing AI and recruitment.

EN



**/ Was können wir gegen
algorithmische Diskriminierung tun?**



/ Sollten wir Algorithmen und KI überhaupt nutzen?

- Ja. So, dass sie uns (allen) nützen.
- Aber: Algorithmen und KI ...
 - ... sind **ein Werkzeug**, das von Menschen entwickelt und eingesetzt wird.
 - ... sind **kein Zweck** an sich.
 - ... sind **keine Naturgewalt**.



/ Mitbestimmungsrecht!

- Welche Entscheidungen sollen automatisiert werden und warum?
- Wer profitiert davon und für wen hat es Nachteile?
- Wer trägt die Verantwortung?
- Gibt es eine Möglichkeit zu widersprechen oder Fragen zu stellen?



/ Lösungsansätze

- Transparenz
- Nachvollziehbarkeit
- Verantwortlichkeit



Wir geben Hilfestellung: Online-Tool zur Folgenabschätzung

FOLGENABSCHÄTZUNGS-TOOL – DE [V6 – current]

Checkliste 1 & Checkliste 2, Stand vom 07.01.2025

Zweck des Tools

Algorithmische und auf Künstlicher Intelligenz (KI) beruhende Systeme haben Auswirkungen auf Menschen und auf Gesellschaften, die ethisch relevant sein können. Deswegen tragen diejenigen, die die Systeme einsetzen, eine besondere Verantwortung. Dieses Instrument zur Folgenabschätzung ist eine Methode, um Anwender*innen zu einem verantwortungsvollen Umgang mit algorithmischen Systemen zu befähigen.

→ **Können wir das Tool frei nutzen?** AlgorithmWatch stellt dieses Impact Assessment Tool zur freien Nutzung zur Verfügung. Wenn Sie sich dabei begleiten lassen möchten oder Fragen haben, wenden Sie sich an info@algorithmwatch.ch (Schweiz) oder info@algorithmwatch.org (Deutschland). Andernfalls freuen wir uns auch über Unterstützung in Form von Spenden – an [AlgorithmWatch CH \(Schweiz\)](#) oder [AlgorithmWatch \(Deutschland\)](#).



DE

[1.1a] Datenschutz: Verarbeitet das algorithmische System sensible Kategorien personenbezogener Daten im Sinne der geltenden Rechtsnormen?

- Ja
 Nein

Hinweis: Reflektieren Sie, welche Daten von Nutzenden in das System eingespeist werden. Können sich darunter "besonders schützenswerte" (gemäss Art. 5c des schweizerischen Datenschutzgesetzes) oder "besondere Kategorien" (gemäss Art. 9 EU-Datenschutzgrundverordnung) personenbezogener Daten befinden?

[1.1b] Könnte der Inhalt, der vom algorithmischen System generiert wurde, bzw. eine Empfehlung, Vorhersage oder Entscheidung, die vom algorithmischen System beeinflusst wurde, Auswirkungen auf die Privatsphäre oder das Privatleben der Menschen, einschliesslich ihres Familienlebens, haben?

- Ja
 Nein

[1.2] Cybersicherheit: a) Haben böswillige Akteure ein besonders grosses Interesse, das algorithmische System zu hacken? b) Könnte das gehackte System genutzt werden, um einen finanziellen Gewinn erzielen – auch durch Erpressung? c) Oder kann ein gehacktes System genutzt werden, um politische Ziele zu erreichen (etwa, um politischen Widerspruch gegen das System zu äussern)?

- Ja
 Nein



EN



/ Ressourcen zu algorithmischer Diskriminierung



EXPLAINER: ALGORITHMISCHE DISKRIMINIERUNG

Wie und warum Algorithmen diskriminieren

Automatisierte Entscheidungssysteme können diskriminieren. Wir erklären, welche Ursachen das hat, was in solchen Fällen passieren kann und warum die bestehenden Gesetze nicht ausreichend vor algorithmischer Diskriminierung schützen.

BLOG 26. JUNI 2024 READ IN ENGLISH #DISKRIMINIERUNG



DE

/ Automatisch aussortiert? Wenn KI-Systeme in Bewerbungsprozessen diskriminieren

KI-unterstützte Systeme in der Personalauswahl können Bewerbungen vorsortieren und bewerten oder Kandidat*innen empfehlen. Solche auf Algorithmen basierende Systeme mögen Personalverantwortlichen Zeit sparen, sie bergen aber auch das Risiko einer systematischen Diskriminierung – die oft unbemerkt bleibt.

BLOG 24. NOVEMBER 2025 #ARBEIT #DISKRIMINIERUNG #EXPLAINER #WORK



DE



30h Course on Fairness in AI Recruitment (OpenCourseWare)

30h Course on Fairness in A...

The **FINDHR (Fairness and Intersectional Non-Discrimination in Human Recommendation)** project aims to develop state-of-the-art anti-discrimination methods, algorithms, and training courses in the field of algorithmic hiring and AI-based human recommendation. One of the project's three core ambitions is to equip professionals, researchers, and practitioners with specialised and advanced skills to **detect, mitigate, and prevent discrimination risks** in algorithmic hiring systems.

To achieve this objective, FINDHR has developed the **30h Course on Fairness in AI Recruitment (OpenCourseWare)**, a comprehensive training programme that combines theoretical foundations, real-world case studies, and hands-on data analysis. The course addresses the technical, legal, and ethical challenges posed by AI-driven recruitment tools, making complex issues of fairness, bias, and intersectional discrimination both accessible and engaging.

The **30h Course on Fairness in AI Recruitment (OpenCourseWare)** is designed to be followed sequentially for maximum learning impact: participants are encouraged to begin with the theoretical modules, continue with the analysis of case studies, and finally apply their knowledge through practical data analysis exercises. This structured approach ensures a deep understanding of discrimination risks and effective strategies for fair and responsible AI deployment in recruitment.

Access credentials

To access the course materials, please use the following credentials:

- **User:** guest_user
- **Password:** Guest123#



EN

POSITIONSPAPIER

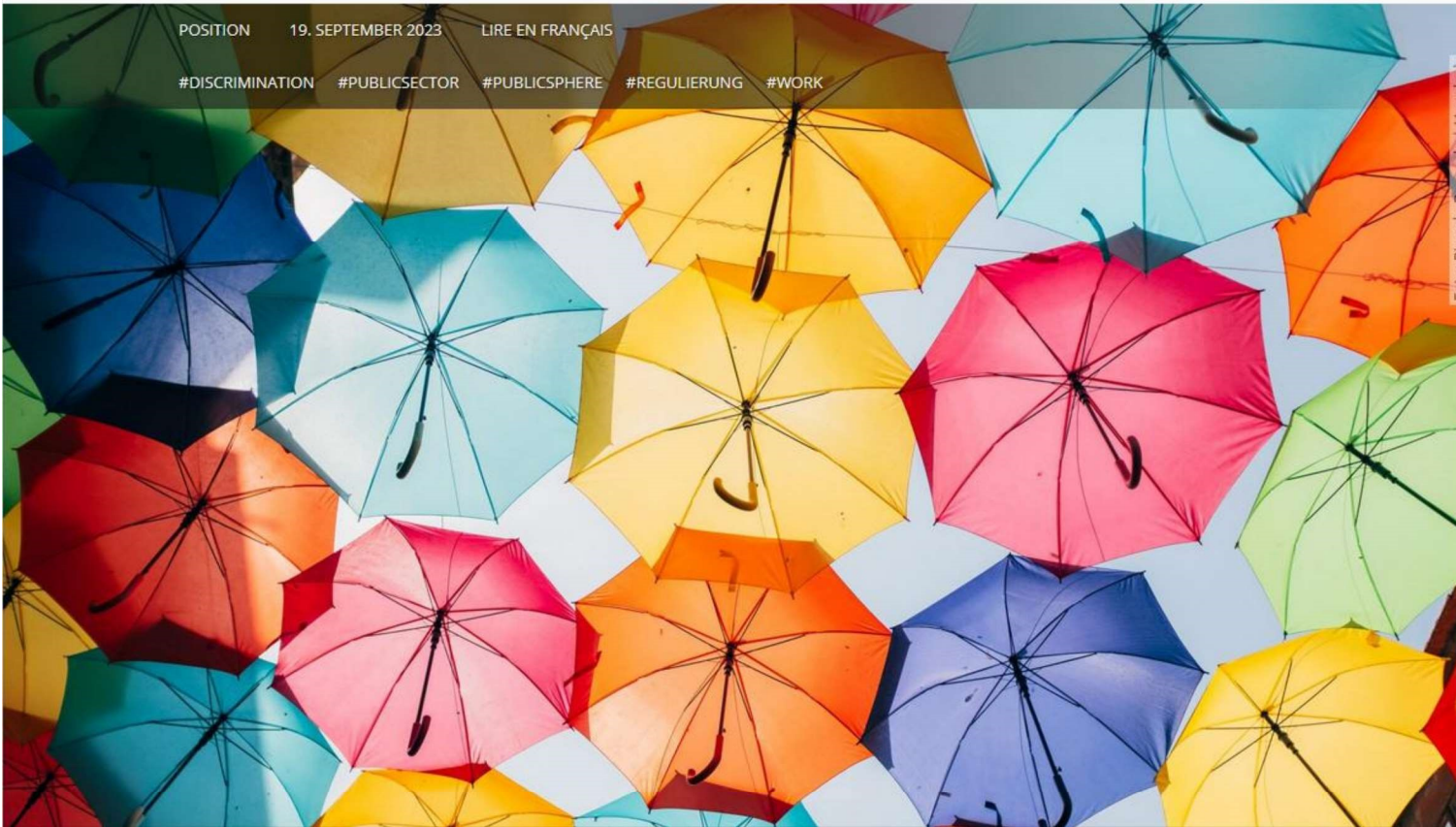
/ Diskriminierende Algorithmen: So gelingt der Schutz

Wenn Algorithmen Prognosen erstellen, Empfehlungen geben oder Entscheidungen fällen, besteht immer das Risiko, dass sie dabei Menschen diskriminieren. Doch wie diskriminieren Algorithmen genau? Und was sollte die Politik dagegen tun? Wir beleuchten den Handlungsbedarf und zeigen Lösungen für einen besseren Schutz vor algorithmischer Diskriminierung.

POSITION 19. SEPTEMBER 2023 LIRE EN FRANÇAIS

#DISCRIMINATION #PUBLICSECTOR #PUBLICSPHERE #REGULIERUNG #WORK

Foto: Ricardo Rieende | Unsplash



DE



/ AutoCheck-Workshops zu automatisierten Entscheidungssystemen und Diskriminierung

Ursachen verstehen, Fälle erkennen, Betroffene unterstützen:
AlgorithmWatch stellt Unterlagen zur Durchführung eines Workshops bereit.

PUBLIKATION 7. SEPTEMBER 2022 READ IN ENGLISH #AUTOCHECK #DISKRIMINIERUNG



DE

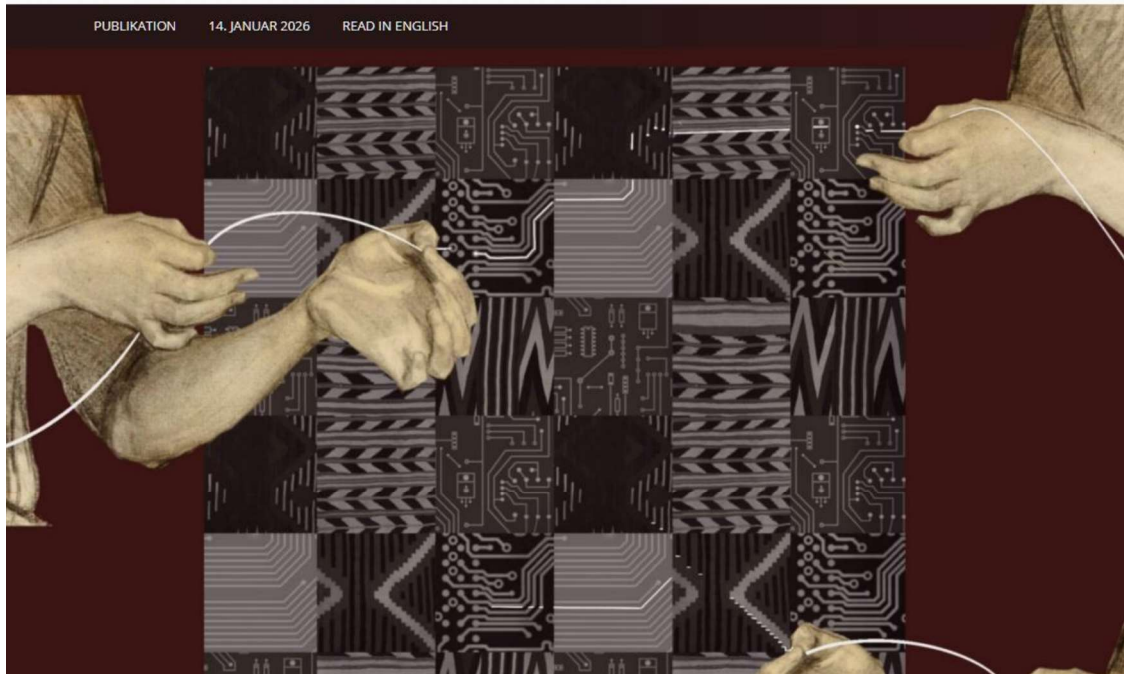


/ Nutzung von KI in der eigenen Organisation: Wie geht AlgorithmWatch damit um?

/ Verantwortungsvolle Nutzung von generativer KI: AlgorithmWatch schlägt diese Richtlinien vor

Ob ChatGPT, Claude oder Gemini, Copilot oder Perplexity – generative KI wirft massive Probleme auf: Viele Ergebnisse sind fehlerhaft und politisch problematisch, der Verbrauch von Energie und Wasser ist enorm. Zugleich sind sie aus dem Alltag nicht mehr wegzudenken. AlgorithmWatch hat eine Richtlinie entwickelt, die dabei hilft, generative KI verantwortungsvoll zu verwenden.

PUBLIKATION 14. JANUAR 2026 READ IN ENGLISH



DE

Sie möchten einen Blick in unseren Leitfaden werfen? Dann können Sie diesen hier anfordern:





/ Unterstützen Sie uns dabei!

Bleiben Sie informiert:

algorithmwatch.org/de/newsletter



Werden Sie Teil unserer Community:

<https://algorithmwatch.org/de/foerdermitglied/>





DANKE FÜR DIE AUFMERKSAMKEIT

/ algorithmwatch.ch | algorithmwatch.org

/ **Lejla Eva Podgoršek** (sie/ ihr)
Senior Policy Manager – Technology & Democracy
podgorsek@algorithmwatch.org
Tel.: +49 30 235 972 423





KI-Agenten Autonomie und Verantwortung

25.03.2026 | Janek Kuberzig | Bundesverband Digitale Wirtschaft (BVDW)

Agenda

1. Was sind KI-Agenten
2. Was bedeutet Ethik bei KI-Agenten
3. KI-Agenten in der Praxis



Was sind KI-Agenten



KI-Agenten | Autonomie und Verantwortung

Was sind KI-Agenten?

	Bot	KI-Assistent	KI-Agent
Zweck	Automatisierung einfacher Aufgaben oder Unterhaltungen	Nutzende bei Aufgaben unterstützen	Aufgaben autonom und proaktiv ausführen
Funktion	Folgt vordefinierten Regeln; eingeschränktes Lernen; grundlegende Interaktionen	Reagiert auf Anfragen oder Prompts, liefert Informationen und führt einfache Aufgaben aus; kann Aktionen empfehlen, aber der Nutzende trifft die Entscheidungen	Kann komplexe, mehrstufige Aktionen ausführen, lernt und passt sich an, kann Entscheidungen unabhängig treffen
Interaktion	Reaktiv; reagiert auf Trigger oder Befehle	Reaktiv; reagiert auf Nutzeranfragen	Proaktiv und zielorientiert

KI-Agenten | Autonomie und Verantwortung

Was sind KI-Agenten?

Autonome Zielverfolgung

Der Agent erhält eine Aufgabe oder formuliert selbstständig ein Ziel basierend auf seiner definierten Rolle.

Planung und Koordination

Er analysiert das Ziel, zerlegt komplexe Aufgaben in Teilaufgaben, priorisiert diese und stimmt sich ggf. mit anderen Agenten ab, insbesondere in Multiagenten-Systemen.

Handlung und Ausführung

Der Agent setzt konkrete Aktionen um, z. B. das Ausfüllen eines Formulars, das Schreiben eines Textes oder das Auslösen automatisierter Prozesse über APIs.

Lernen und Anpassung

Auf Basis von Rückmeldungen aus der Umwelt optimiert der Agent sein Verhalten. Er lernt aus Fehlern oder Erfolgen und passt seine Strategie für künftige Aufgaben an.



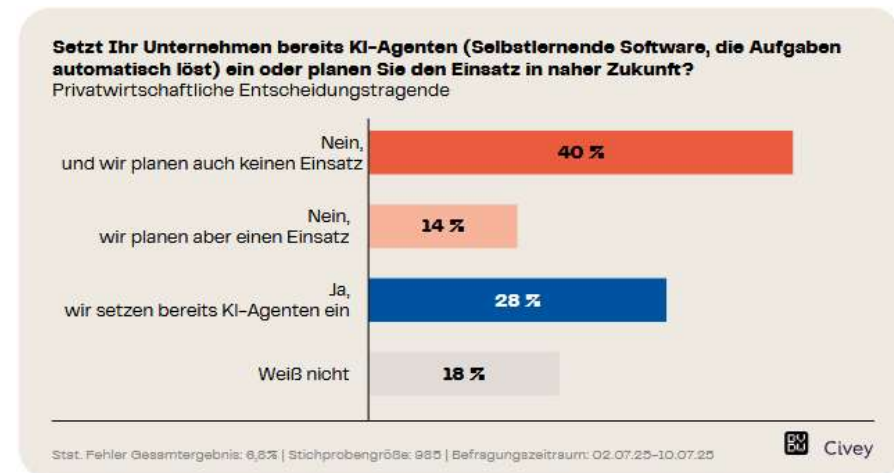
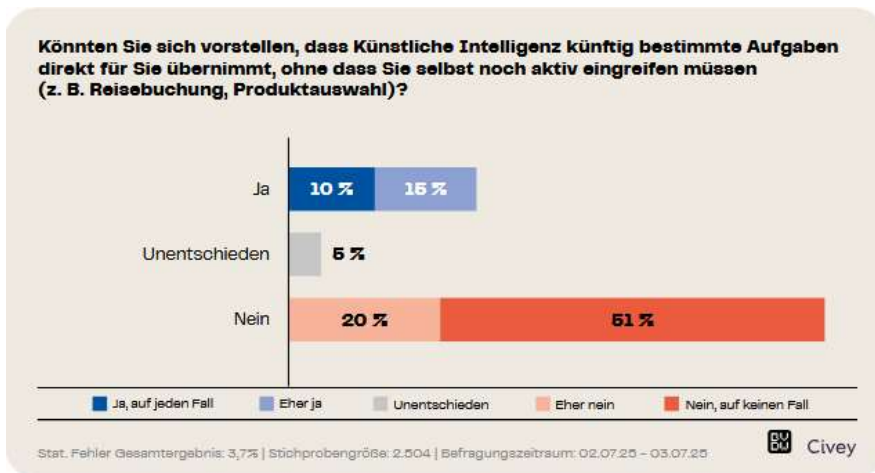
Was bedeutet Ethik bei KI-Agenten



KI-Agenten | Autonomie und Verantwortung

Was bedeutet Ethik bei KI-Agenten

Weder eine Mehrheit der **Verbraucher*innen** noch der **Unternehmen** nutzen zurzeit KI-Agenten.



Wir haben auf beiden Seiten ein Vertrauensproblem im Markt.

KI-Agenten | Autonomie und Verantwortung

Was bedeutet Ethik bei KI-Agenten

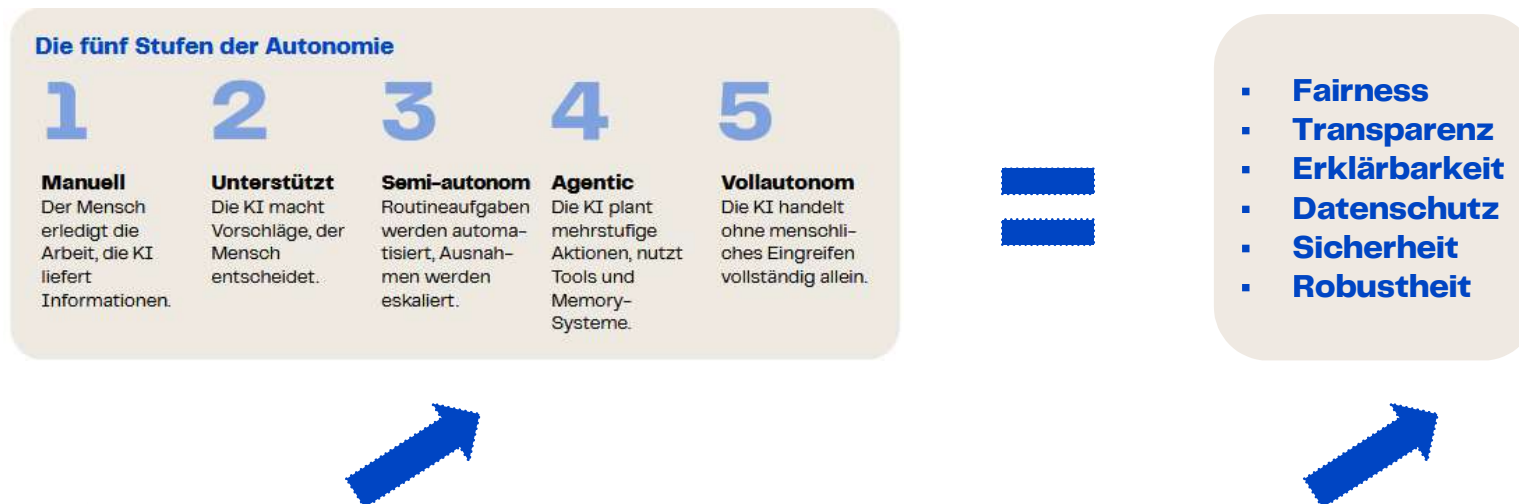
Sechs Prinzipien für eine ethische KI

- **Fairness:** KI-Systeme sollen niemanden diskriminieren oder benachteiligen.
- **Transparenz:** Die Funktionsweise von KI-Systemen soll einsehbar sein.
- **Erklärbarkeit:** KI-Entscheidungen sollen nachvollziehbar sein.
- **Datenschutz:** Der Schutz personenbezogener Daten soll gewährleistet sein.
- **Sicherheit:** Fehlfunktionen, Manipulationen und Missbrauch sollen verhindert werden.
- **Robustheit:** KI-Systeme sollen auch unter unsicheren Bedingungen zuverlässig funktionieren.

KI-Agenten | Autonomie und Verantwortung

Was bedeutet Ethik bei KI-Agenten

Je höher der Autonomiegrad einer KI, desto höher die ethischen Anforderungen an ihren Einsatz.



KI-Agenten | Autonomie und Verantwortung

Was bedeutet Ethik bei KI-Agenten

Autonomiegrad und Anwendungsfall des KI-Agenten definieren menschliche Aufsicht.

Grad der Aufsicht

- **Human-in-the-Loop (HITL):** Der Mensch muss kritische Aktionen freigeben.
- **Human-on-the-Loop (HOTL):** Der Mensch überwacht und kann schnell eingreifen.
- **Human-in-Command (HIC):** Der Mensch setzt Ziele und Vorgaben, das System unterstützt die Ausführung.



Autonomiestufe zuweisen

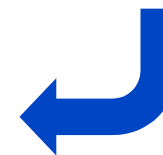


Den Worst-Case-Schaden bestimmen/schätzen

Hoch = HITL

Mittel = HOTL

Niedrig = HIC





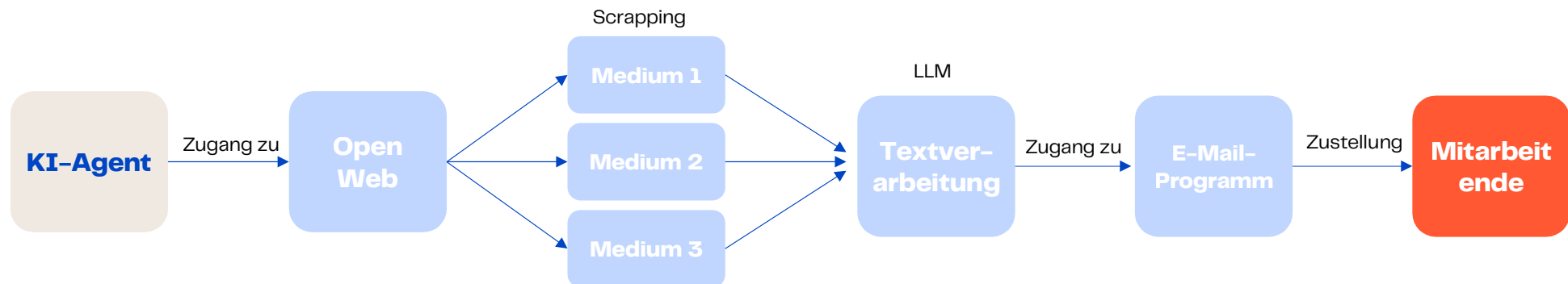
KI-Agenten in der Praxis



KI-Agenten | Autonomie und Verantwortung

KI-Agenten in der Praxis

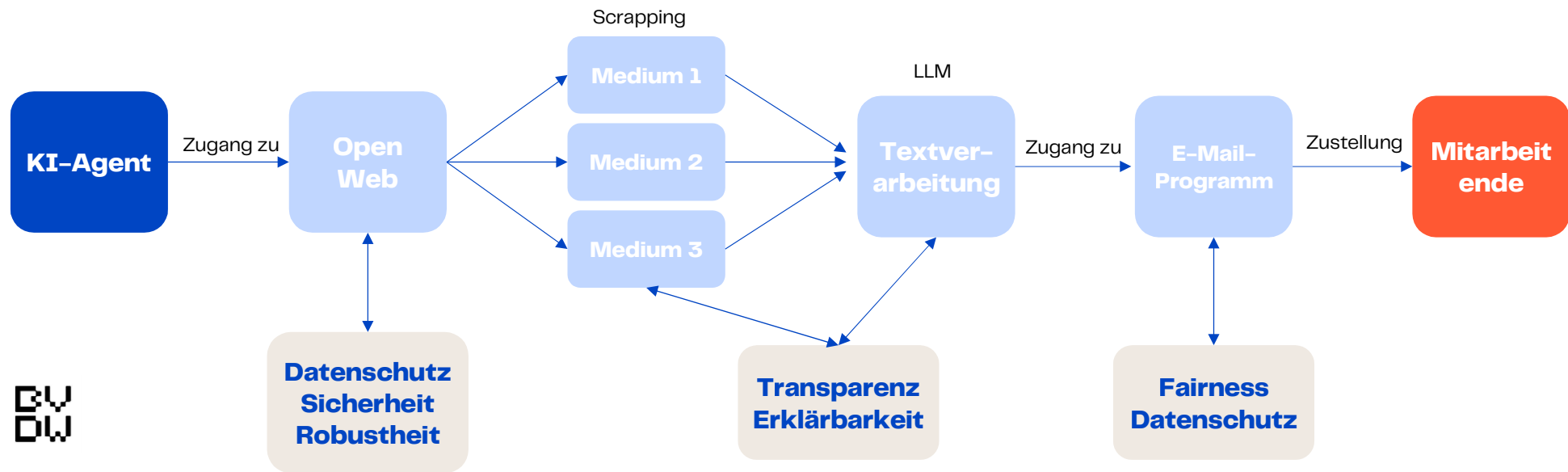
Use Case: Ein KI-Agent bekommt das Ziel täglich ein Pressescreening mit vorher definierten Medien und Themen durchzuführen. Anschließend soll ein Briefing erstellt werden und an eine Gruppe von Mitarbeitenden verschickt werden.



KI-Agenten | Autonomie und Verantwortung

KI-Agenten in der Praxis

Use Case: Ein KI-Agent bekommt das Ziel täglich ein Pressescreening mit vorher definierten Medien und Themen durchzuführen. Anschließend soll ein Briefing erstellt werden und an eine Gruppe von Mitarbeitenden verschickt werden.



KI-Agenten | Autonomie und Verantwortung

KI-Agenten in der Praxis

Use Case: Ein **multimodales KI-Agenten-System** erhält das Ziel, Kundenanfragen bei technischen Störungen effizient und koordiniert zu bearbeiten. Ein **Kundenservice-Agent** verarbeitet eingehende Anfragen über Text, Sprache und Bilder (z. B. Schadensfotos) und klassifiziert das Anliegen nach Art und Dringlichkeit.

Anschließend übergibt er strukturierte Informationen an einen **Produktions-Agent**, der Maschinendaten, Sensorwerte und Wartungshistorien analysiert und bewertet, ob ein Produktionsrisiko oder ein drohender Stillstand vorliegt.

Falls erforderlich, kommuniziert der Produktions-Agent mit einem **Beschaffungs-Agent**, der automatisch Ersatzteilverfügbarkeiten prüft, Lieferzeiten bewertet und Beschaffungsvorgänge priorisiert oder anstößt.

Ein übergeordneter **Governance-Agent** überwacht die Entscheidungen der beteiligten Agenten, prüft definierte Schwellenwerte (z. B. Kosten, Risiko, Auswirkungen auf Kunden) und entscheidet, wann eine menschliche Aufsicht eingebunden oder ein Prozess gestoppt werden muss.

KI-Agenten | Autonomie und Verantwortung

KI-Agenten in der Praxis

AGENT	ZENTRALE AUFGABE	GENUTZTE DATEN	GOVERNANCE-RELEVANZ
Kundenservice-Agent	Erfassung und Klassifikation von Kundenanfragen	Text, Sprache, Bilder, Videos	Fairness, Datenschutz, Erklärbarkeit
Produktions-Agent	Bewertung von Produktionsrisiken und Maßnahmen	Sensorik, Maschinendaten, Wartungshistorien	Erklärbarkeit, Robustheit, Sicherheit
Beschaffungs-Agent	Prüfung und Auslösung von Beschaffungsvorgängen	Lagerbestände, Lieferzeiten, Kosten	Transparenz, Robustheit, Sicherheit
Governance-Agent	Überwachung, Eskalation und Regelprüfung	Entscheidungsprotokolle aller Agenten	Transparenz, Erklärbarkeit, Sicherheit
Menschliche Aufsicht	Zielvorgabe, Freigabe und Haftung	Aggregierte Entscheidungsinformationen	Rechtliche und ethische Gesamtverantwortung



THANK YOU!

Kontaktieren Sie mich gerne!



Janek Kuberzig

**Public Affairs Manager | Data & Technology
Bundesverband Digitale Wirtschaft (BVDW)**

Sie erreichen mich unter:

E-Mail: kuberzig@bvdw.org

LinkedIn: www.linkedin.com/in/janek-kuberzig

Weitere Informationen:

www.bvdw.org

Paper zu KI-Agenten:



Paper zu Verantwortung bei KI-Agenten:

