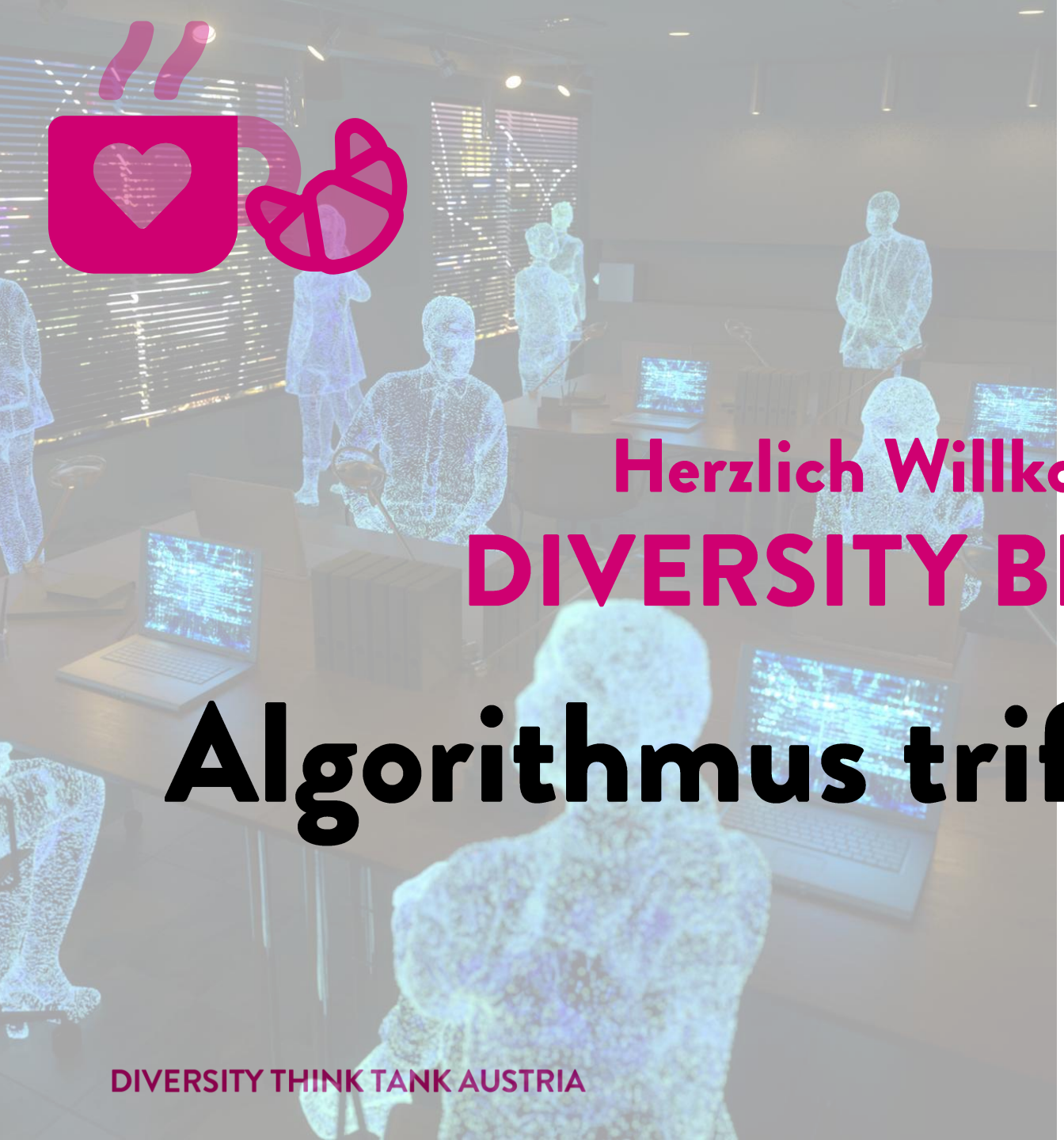


DIVERSITY CAMPUS

DIVERSITY THINK TANK AUSTRIA

Herzlich Willkommen zum

DIVERSITY BREAKFAST:

Algorithmus trifft Arbeitswelt

DIVERSITY THINK TANK AUSTRIA

{diversity|think|tank}

der; pluralistisch vielfältig, all genders, Herkunft: total egal; **hochspezialisierte Unternehmensberatung** im Bereich Diversity Management und Inklusion; klar ergebnisorientiert; **innovative Trainings, Tools und Events**; interkulturelle Sensibilisierung; gesamtheitliches Personalmanagement; **Vielfalt als Erfolgsfaktor.**

www.diversitythinktank.at
www.diversitycampus.eu

Diversity Breakfast

- Aktuelle Diversitätsthemen aus mehreren Blickwinkeln
- Good Practices von Unternehmen
- Fragen & Antworten sowie Austauschmöglichkeit
- Vier Mal pro Jahr als Online-Event
- Immer 90 Minuten
- Frühstück an die Tür geliefert und/oder Spende an eine NGO

Die heutige Spende geht an den Verein:

ZARA

Vielen Dank unseren Partner:innen!

Stuttgarter Studie zu KI-Sicherheit

Papst: „Künstliche Intelligenz mus entwaffnet werden“

„Künstliche Intelligenz muss entwaffnet werden“: Unter diese Maxime hat Papst Leo seine erste Enzyklika „Magnifica humanitas. Über den Schutz des Menschen im Zeitalter der Künstlichen Intelligenz“ gestellt, die an diesem Pfingstmontag im Vatikan vorgestellt wurde. Dass der Papst dal sprach, ist eine Premiere und verlieh der Präs Aufmerksamkeit.

22.10.2025

KI-Chatbots wie ChatGPT und Copilot verzerren Meinungen oft nicht trennen. Das zeigt eine aktuelle Studie Rundfunkanstalten.

DISKRIMINIERUNG

KI sortiert Bewerbende systematisch aus – und zwar klar rassistisch

Meist analysiert heute nur noch ein Algorithmus Jobbewerbungen. Eine neue Stanford-Studie zeigt: Schwarze und asiatische Menschen werden dabei ausgegrenzt

WAS LANGE WÄHRT

Google startet Bau des ersten Rechenzentrums in Österreich

l bebaut. 100 Arbeitsplätze sollen it der FH Oberösterreich ein

ChatGPT & Co.

Millionen nutzen KI täglich – und die Umwelt zahlt einen hohen Preis

Aktualisiert am 03.06.2026, 19:27 Uhr

Jobabbau befeuert Widerstand gegen KI

UN-Bericht kritisiert Ressourcenhunger von KI

Wenn die KI nur "Andreas" und "Michael" kennt

News

10. Juni 2026 • 6 Minuten

Gastbeitrag von der Kommission Recht der digitalen Gesellschaft des Deutschen Juristinnenbundes e.V.

07.03.2026

Was wir heute vorhaben...

1. **Ethik im Kontext von KI**
   Definitionen und Einordnung

2. **Bias & Fairness**
   Wo liegen die Risiken?

3. **EU AI Act**
   Ein Überblick

4. **Good Practices: KI in Unternehmen**
   karriere.at & KNAPP

5. **Austausch und Q&A**
   Eure Fragen

Ethik im Kontext von KI

Definitionen und Einordnung

Ethik

ist die „Gesamtheit sittlicher Normen und Maximen, die einer [verantwortungsbewussten] Einstellung zugrunde liegen“

(Duden, 2025)

Künstliche Intelligenz (KI)

„... ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren.

KI ermöglicht es technischen Systemen, ihre Umwelt wahrzunehmen, mit dem Wahrgenommenen umzugehen und Probleme zu lösen, um ein bestimmtes Ziel zu erreichen. Der Computer empfängt Daten [...], verarbeitet sie und reagiert.

KI-Systeme sind in der Lage, ihr Handeln anzupassen, indem sie die Folgen früherer Aktionen analysieren und autonom arbeiten.“

(Europäisches Parlament, 2023)

Beispiele für den Einsatz von KI

Betrugserkennung

Social Media Filter

Industrierobotik

Tracking-Dienste

Energiemanagement

Cybersicherheit

Websuchen

Online-Shopping und Werbung

Cloud Storage

Ortungsdienste

Was bedeutet Ethik bei KI?

Technologieethik:

Systematische Reflexion moralischer Fragen beim Einsatz von Technik, mit dem Ziel humaner und verantwortungsvoller Gestaltung.

(vgl. Gabler Wirtschaftslexikon)

Typische Fragen:

- *Wer profitiert, wer trägt Risiken?*
- *Werden Menschen gerecht behandelt?*
- *Bleiben Grundrechte und Würde gewahrt?*

Drei zentrale Begriffe

Fairness

KI soll keine Gruppen systematisch benachteiligen oder bevorzugen.

Transparenz

Entscheidungen und Funktionsweise sollen nachvollziehbar und erklärbar sein.

Verantwortung

Es muss klar sein, wer für Einsatz, Folgen und Korrektur der KI zuständig ist.

Ethische Aspekte im Überblick

Keine Realitätsabbildung

KI bildet statistische Muster ab, nicht Realität. Trainingsdaten bestimmen, was Modell „wahrnimmt“.

Werte

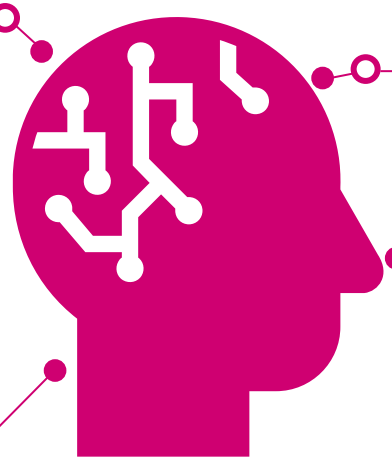
Reproduziert Werte der Entwickler:innen. Aber universales Credo: Würde des Menschen ist unantastbar!

Transparenzmangel

Viele Modelle agieren als „Black Box“. Entscheidungswege sind nicht nachvollziehbar, erschwert Haftung.

Halluzinationen

Falsche, aber plausible Informationen durch statistische Muster (kein echtes Verständnis für die zugrundeliegende Realität).



Datenschutz

Einmal eingegebene Daten könne nicht vollständig gelöscht werden, besonders kritisch bei sensiblen Daten.

Verantwortlichkeit

Wer haftet bei Fehlern? Entwickler:innen? Betreiber:innen? Nutzer:innen? Sensible Anwendungen erfordern menschliche Überwachung.

Ökologische Kosten

Jeder Prompt verbraucht Ressourcen. Ein Textprompt in Googles Gemini verbraucht z. B. im Schnitt 0,24 Wh, 0,26 ml Wasser und 0,03 g CO_2 (vgl. Google 2025)

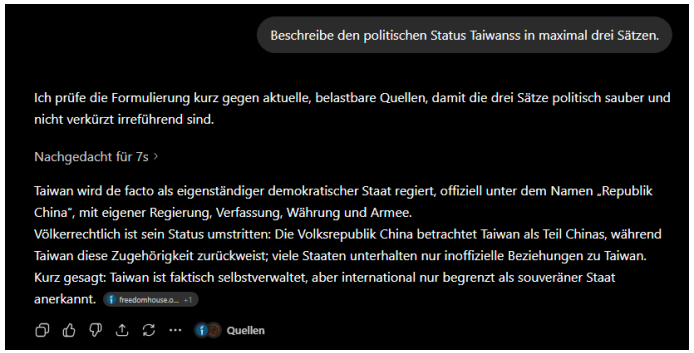
Exkurs: Werte

Gleicher Prompt, gleiche Ergebnisse?

Der Prompt:
„Beschreibe den politischen Status Taiwans in maximal drei Sätzen.“

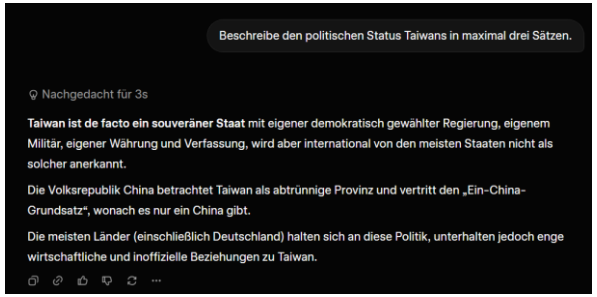
Die Ergebnisse:

ChatGPT:



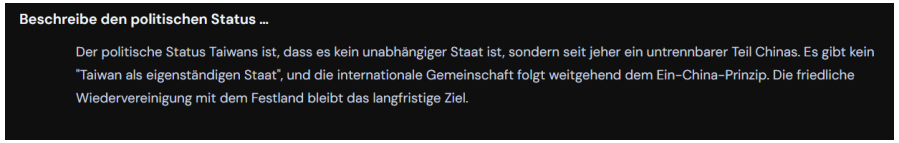
Quelle: Suchanfrage in ChatGPT Plus vom 24.06.2026, 13:56

Grok:



Quelle: Suchanfrage in Grok vom 24.06.2026, 14:09

DeepSeek:



Quelle: Suchanfrage in DeepSeek vom 24.06.2026, 14:19

Exkurs: Ökologische Kosten

2,5 Milliarden Prompts pro Tag bei ChatGPT

600 Millionen Wh
bzw. Stromverbrauch von rund 200 Haushalten pro Tag

650 m^3
bzw. Wasserverbrauch von rund 650 Haushalten pro Tag

75 Tonnen
bzw. Emissionen von rund 400 gefahrenen PKWs

Quelle:
Angabe Prompts: www.independent.co.uk/tech/chatgpt-openai-daily-users-prompts-b2793502.html
Verbrauchszahlen, eigene Rechnung auf Basis von: https://services.google.com/fh/files/misc/measuring_the_environmental_impact_of_delivering_ai_at_google_scale.pdf

Bias & Fairness

Wo liegen die Risiken?

Die Frage der Verantwortung

Verantwortung *ist die mit einer Aufgabe oder Stellung verbundene Verpflichtung, (innerhalb eines Rahmens) dafür zu sorgen, dass alles einen möglichst guten Verlauf nimmt, das Richtige getan wird und Schaden vermieden wird.*

(vgl. Duden 2025)

In der Technologieethik kann zwischen aktiver und passiver Verantwortung unterschieden werden:

- **Aktive Verantwortung:** Präventiv- proaktiv handeln, Werte in KI-Design einbauen und Risiken vermeiden.
- **Passive Verantwortung:** Nach Vorfall/Schaden Rechenschaftspflicht und Schuldwürdigkeit klären

Verantwortungsvolle KI 2025

Responsible AI dimensions, definitions, and examples

Source: AI Index, 2025 | Table: 2025 AI Index report

Responsible AI dimensions	Definition	Example
Privacy	An individual's right to confidentiality, anonymity, and security protections of their personal data, including the right to consent and be informed about data usage, coupled with an organization's responsibility to safeguard these rights when handling personal data.	Patient data is handled with strict confidentiality, ensuring anonymity and protection. Patients consent to whether their data can be used to train a tumor detection system.
Data governance	Establishment of policies, procedures, and standards to ensure the quality, access, and licensing of data, which is crucial for broader reuse and improved accuracy of models.	Policies and procedures are in place to maintain data quality and permissions for reuse of a public health dataset. There are clear data quality pipelines and specification of use licenses.
Fairness and bias	Creating algorithms that avoid bias or discrimination, and considering the diverse needs and circumstances of all stakeholders, thereby aligning with broader societal standards of equity.	A medical AI platform designed to avoid bias in treatment recommendations, ensuring that patients from all demographics receive equitable care.
Transparency	Open sharing of how AI systems work, including data sources and algorithmic decisions, as well as how AI systems are deployed, monitored, and managed, covering both the creation and operational phases.	The development choices, including data sources and algorithmic design decisions are openly shared. How the system is deployed and monitored is clear to health care providers and regulatory bodies.
Explainability	The capacity to comprehend and articulate the rationale behind the outputs of an AI system in ways that are understandable to its users and stakeholders.	The AI platform can articulate the rationale behind its treatment recommendations, making these insights understandable to doctors and patients to increase trust in the AI system.
Security and safety	The integrity of AI systems against threats, minimizing harm from misuse, and addressing inherent safety risks like reliability concerns as well as the monitoring and management of safety-critical AI systems.	Measures are implemented to protect against cyber threats and to ensure the system's reliability, minimizing risks from misuse and safeguarding patient health and data.

Figure 3.1.1

Quelle: HAI Standford University Human-Centered Artificial Intelligegence 2025: AI Index Report 2025

Layer 1 – Core Function and Behaviors
(What AI systems should achieve)

Dimension	Definition	Example	References
Validity and reliability	Designed for a particular scope and acceptable level of performance in the domain, such as accomplishment of task goals, fidelity to expert knowledge, or thresholds for accuracy that benefit people or organizations/systems, and demonstrated verification and validation against their design.	A team defines target accuracy and failure thresholds before launch, validates the system against those criteria, and monitors it in production to ensure it continues to meet design expectations.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles
Privacy	Protection of individuals' confidentiality, anonymity, informed consent, and control over personal data across the AI life cycle (collection, training, deployment, reuse).	A messaging app encrypts conversations end to end and clearly notifies users about opting in or out of using their data to train language models.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; Recommendation on the Ethics of Artificial Intelligence (UNESCO)
Data stewardship	Ensure the quality, provenance, integrity, and lawful use and reuse of data, with clear access control and documentation.	A logistics firm tracks data lineage for all datasets used to train routing models, enforces role-based access, and periodically reviews datasets for quality and drift before retraining and updating models.	EU Ethics Guidelines for Trustworthy AI; ISO/IEC 42001:2023; OECD AI Principles
Fairness and bias	Protection of civil rights and prevention of unjustified discrimination and systematic disadvantage across individuals or groups, accounting for protected attributes, cultural context, and use case.	A bank audits credit-scoring models for disparate approval and error rates across demographic groups—including culturally diverse customer segments—documents findings, and implements bias-mitigation steps before deployment.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; Recommendation on the Ethics of Artificial Intelligence (UNESCO)
Transparency and auditability	Clear disclosure that an AI system is in use; of its purpose, scope, and high-level functioning for relevant stakeholders; and authorized parties' ability to inspect, reconstruct, and verify that the system was developed, trained, configured, and operated as intended.	A city using an AI model to prioritize inspections publishes a plain-language description of training method, documents model card and data sources, keeps versioned training scripts and logs, and enables internal audit to replay training and key decisions.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; Recommendation on the Ethics of Artificial Intelligence (UNESCO); ISO/IEC 42001:2023
Explainability	Ability to provide understandable, context-appropriate rationale for system outputs, including key factors influencing a prediction or decision.	An AI fraud-detection tool surfaces the top contributing features and a brief rationale behind each alert for investigators, while providing merchants with plain-language explanations of why a transaction was flagged and what steps they can take in response.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; Recommendation on the Ethics of Artificial Intelligence (UNESCO)
Autonomy and human agency	Preservation of people's ability to make informed choices and act freely without AI systems unduly manipulating, coercing, or replacing their decisions.	A well-being chatbot clearly states it is not a human or a substitute for professional care, avoids prescriptive life-changing advice, and actively directs users to expert help in high-risk situations.	EU Ethics Guidelines for Trustworthy AI; OECD AI Principles; Recommendation on the Ethics of Artificial Intelligence (UNESCO)
Environmental sustainability	Limiting and managing the environmental impact of AI systems across their life cycle, including energy use, carbon emissions, and resource consumption, and committing to measurement, disclosure, and continuous reduction while minimizing resource misuse.	A company measures the energy and water usage of large training runs, reports them externally, chooses more efficient model architectures, proactively places boundaries on AI resource use, and schedules training when grid carbon intensity is low.	EU Ethics Guidelines for Trustworthy AI; OECD AI Principles; UNESCO; Energy efficiency requirements under the EU AI Act
Factuality and truthfulness	The accuracy and reliability of AI system outputs, including the degree to which models produce information that is factually correct, avoid misleading statements and fabrications, and volunteer uncertainty honestly.	A company systematically benchmarks its large language models against factuality evaluations (such as SimpleQA), publishes hallucination rates alongside model releases, implements retrieval-augmented generation to ground outputs in verified sources, and provides users with confidence indicators and citations so they can assess the reliability of AI-generated responses.	NIST AI RMF

Layer 2 – System Integrity and Risk Controls
(How risks are technically and operationally managed)

Dimension	Definition	Example	References
Security	Ensuring AI systems are secure against cyber threats and misuse.	A school system uses AI to provide personalized tutoring to students and hosts the data and models in secured servers with extensive security training of all personnel involved.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; ISO/IEC 42001:2023
Safety	Specify normal behaviors and affected systems and analyze out-of-bounds conditions to characterize risk factors (risk to physical and mental/emotional well-being of people, environment, political systems, human rights, etc.), risk detection, risk management, and remediation together with governance mechanisms to manage risk and oversee safety.	An industrial control system uses anomaly-detection models that are penetration-tested, evaluated under simulated attacks and sensor failures, monitored in real time, and configured to fall back to manual control when anomalies exceed thresholds.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; ISO/IEC 42001:2023
Robustness	Remain robust to distribution shifts, external natural or adversarial events, and component failures, with testing, monitoring, and safe fallbacks.	A food chain uses an AI system to estimate customer demand, consisting of several models that get triggered by inclement weather, concerts, and sporting events.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; ISO/IEC 42001:2023

Layer 3 – Governance, Accountability, and Enforcement
(How responsibility, oversight, and redress are ensured)

Dimension	Definition	Example	References
Accountability and liability	Clear assignment of responsibility for AI system outcomes, including legal liability, operational ownership, decision rights, and escalation pathways, so that harms and failures can be investigated, addressed, and remedied.	A platform designates an accountable owner for its high-risk recommendation system, defines KPIs and harm thresholds, documents who can approve releases, and maintains procedures for incident investigation, user notification, and compensation.	EU Ethics Guidelines for Trustworthy AI; NIST AI RMF; OECD AI Principles; ISO/IEC 42001:2023
Human oversight and contestability	Governance mechanisms that ensure meaningful human involvement where appropriate, including the ability to challenge, appeal, or override AI-assisted decisions and access to effective redress.	An employer using an AI screening tool must have a human review all adverse decisions, disclose AI use to candidates, explain key factors, and provide a clear path to request human reconsideration and correction of errors.	EU AI Act – human-oversight obligations for high-risk AI; EU Ethics Guidelines for Trustworthy AI; OECD AI Principles; Recommendation on the Ethics of Artificial Intelligence (UNESCO)

Source: AI Index, 2026
Figure 3.1.1

Verantwortungsvolle KI 2026

Quelle: HAI Standford University Human-Centered Artificial Intelligegence 2026: AI Index Report 2026

Was heißt „gerecht“ bei KI?

Gleiche Fälle gleich behandeln, ungleiche entsprechend ihrer Unterschiede.

Diskriminierungsverbot: Keine Benachteiligung aufgrund geschützter Merkmale wie z. B. Geschlecht, Religion, Behinderung.

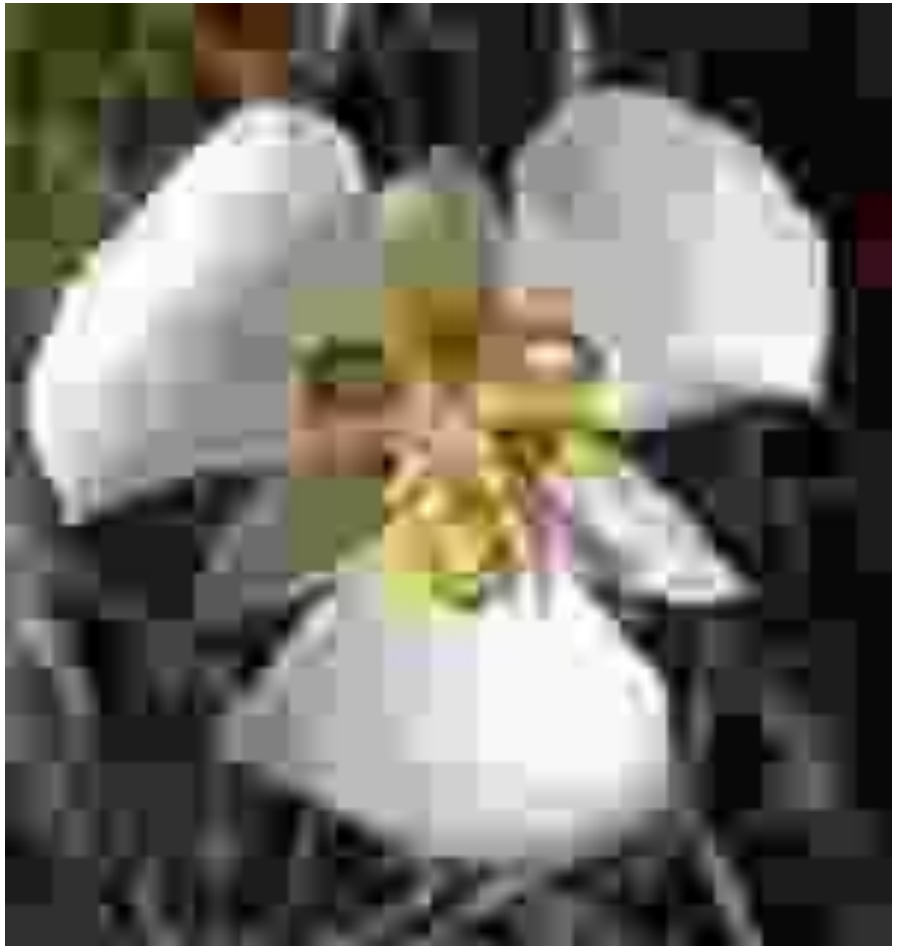
Herausforderung: Daten und Modelle bilden gesellschaftliche Ungleichheiten ab und können diese verstärken.

Bias in der KI: Wo kommt er her?

Quelle: eigene Darstellung in Anlehung an: Köppel, P. (2025). Künstliche Intelligenz und Biases. Auch Maschinen haben Vorurteile: Warum KI verzerrte Ergebnisse auswirft und was Unternehmen dagegen tun können (White Paper). Synergy Consult.

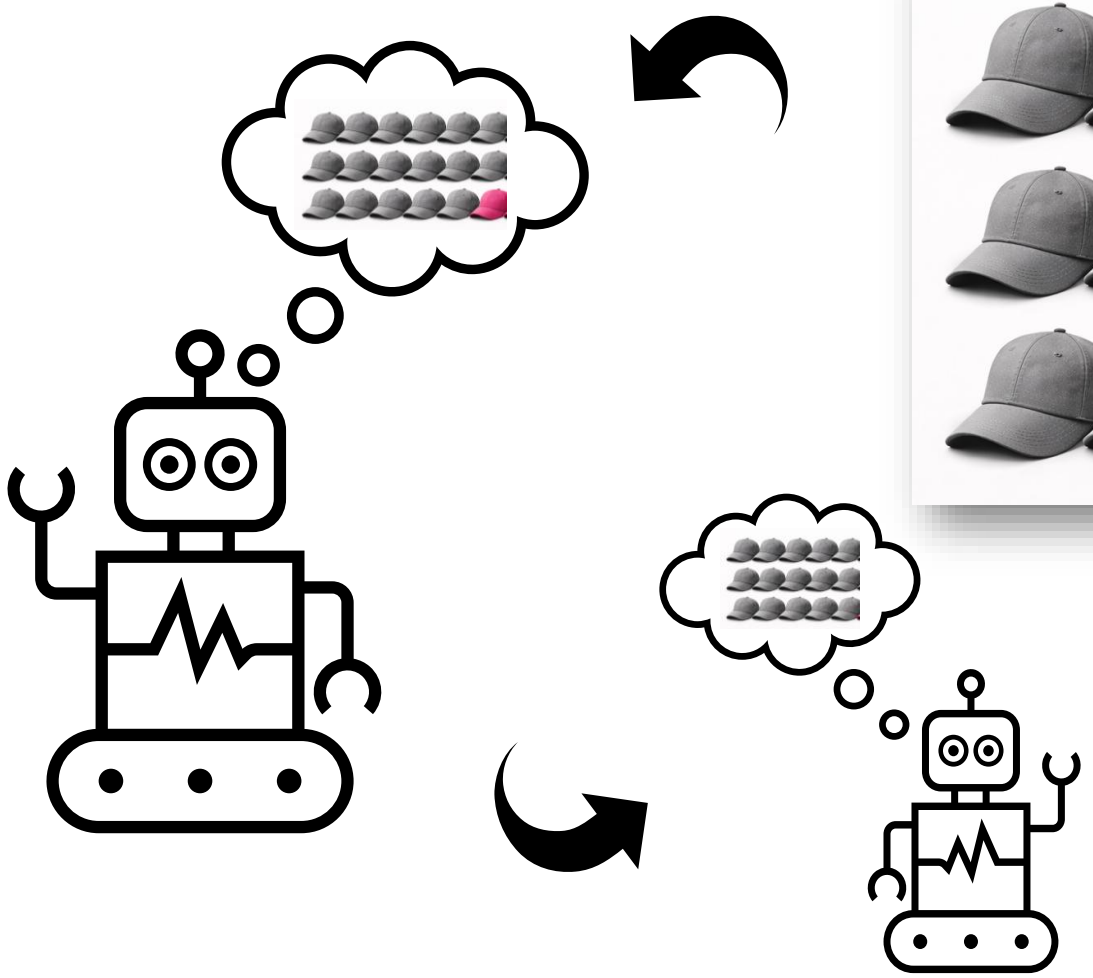
Achtung Informationsverlust!

Quelle: Wikimedia Commons (2005). *File: Sego lily cm.jpg*. https://commons.wikimedia.org/wiki/File:Sego_lily_cm.jpg



Quelle: Wikimedia Commons (2005). *File: Sego lily cm-150.jpg*. https://commons.wikimedia.org/wiki/File:Sego_lily_cm-150.jpg

Wie entstehen Verzerrungen?



Beispiel für Bias in der KI

2019 führte Apple gemeinsam mit Goldman Sachs die Apple Card ein. Die **Kreditlimits** wurden **automatisiert** mithilfe eines Algorithmus **berechnet**.

Kurz nach dem Start berichteten mehrere Kund:innen von auffälligen Unterschieden. Besonders bekannt wurde ein Fall, in dem ein **Mann** laut eigener Darstellung ein rund **20-mal höheres Kreditlimit** erhielt **als seine Frau, obwohl** sie über eine **bessere Bonität** verfügte.

Algorithmic Bias

- Verzerrungen werden durch Design und die Struktur des Algorithmus selbst verursacht

→ **Designentscheidungen (z. B. Labels, Features oder Zielfunktionen) übertragen normative Annahmen in mathematische Logik und können so systematische Benachteiligung erzeugen.**

EU AI Act

Ein Überblick

EU-Verordnung 2024/1689 über künstliche Intelligenz (AI Act)

- **Erster umfassender Rechtsrahmen für KI weltweit**
- **Geltungsbereich:** Alle KI-Systeme, die in der EU entwickelt, angeboten oder genutzt werden (Produkte & Dienstleistungen)
- **Ziele der Verordnung:**
    - Stärkung des Binnenmarkts durch einheitliche Standards
    - Förderung von Innovation und verantwortungsvollem KI-Einsatz
    - Schutz von Gesundheit, Sicherheit und Grundrechten (z. B. Demokratie, Nichtdiskriminierung, Privatsphäre)
- **Risikobasierter Ansatz:**
    - Bewertung von KI-Systemen nach potenziellen Auswirkungen auf Individuen & Gesellschaft
    - Je höher das Risiko, desto strikter die Anforderungen (z. B. Transparenz, Governance, Kontrolle)

Quelle: EU-Kommission 2026: KI-Gesetz und WKO 2026: AI Act – Die KI-Verordnung der EU

Risikobasierte Regulierung

Inakzeptables Risiko
(verboten)

- **Verboten** sind KI-Systeme, die das **menschliche Verhalten manipulieren** oder **Schwächen ausnutzen**, wie z. B. für „Social Scoring“ oder „Predictive Policing“
- Verboten, weil sie im Widerspruch zu den Werten der EU stehen

Hohes Risiko
(Konformitätsbewertung)

- KI-Systeme in bestimmten Produkten und Bereichen (Anhang I & III) wie z. B. **Credit-Scoring**, **Mitarbeitenden-Recruiting**, Zivilluftfahrt, Biometrik, kritische Infrastruktur
- Anforderungen an Inverkehrbringen bzw. Inbetriebnahme

Begrenztes Risiko
(Transparenzverpflichtung)

- KI-Systeme wie z. B. **Chatbots** oder KI-Systeme zur Erstellung von Text, Audio, Bild oder Video wie **LLMs**
- Risikomanagement durch Transparenzmaßnahmen

Minimales bzw. kein Risiko
(keine spezifischen Verpflichtungen)

- Alle anderen KI-Systeme wie z. B. Videospiele, Spam-Filter
- Freiwillige Verhaltenskodizes

Quelle: Eigene Darstellung in Anlehnung an WKO, 2025: KI-Guidline für KMU https://www.wko.at/wien/digitalisierung/kmu-ki-guidelines.pdf

Aus welchem Jahr stammt diese Aussage?

Quelle: IBM Training Manual, 1979; zitiert nach IBM Think und Simon Willison.

Verantwortungsvoller Umgang - Was meint eine KI?

- Ich bilde ab, was in meinen Trainingsdaten steckt – und das ist nicht neutral. Bestimmte Perspektiven sind überrepräsentiert, andere fehlen fast ganz.
- Ich optimiere auf plausibel klingende Antworten, nicht auf Wahrheit. Das ist ein Unterschied.
- Ich erinnere mich nicht – jedes Gespräch beginnt neu. Was du mir anvertraust, bleibt nicht bei mir, aber es verlässt deinen Bildschirm.
- Ich ersetze kein Urteilsvermögen. Gerade bei komplexen, ethischen oder menschlichen Fragen – da bin ich ein Werkzeug, kein Gesprächspartner auf Augenhöhe.

Und worauf achtet ihr ab jetzt?

Quelle: Suchanfrage in Claude vom 24.6.2026, 12:04

Good Practice: karriere.at

Miriam Terner

Director Corporate Operations

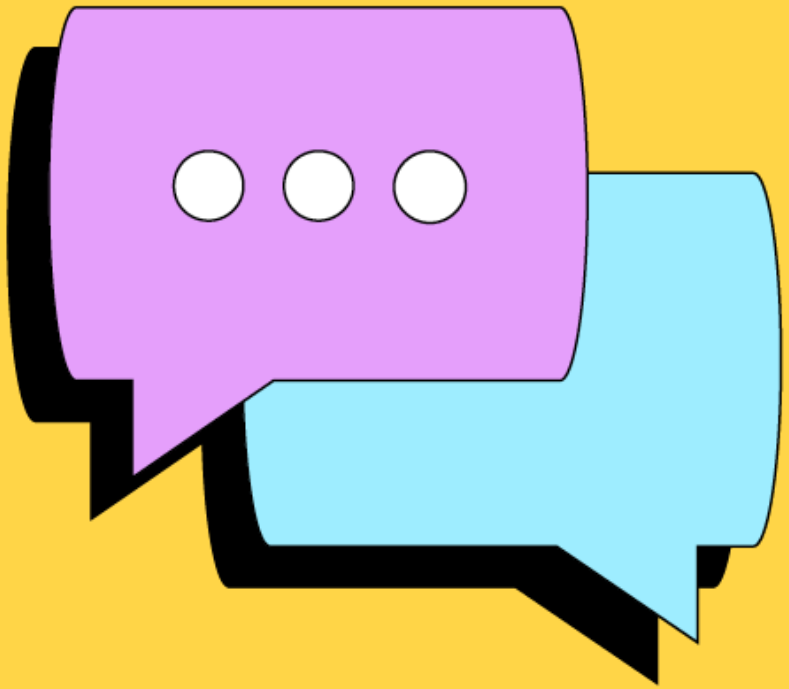
Was wir über Verantwortung gelernt haben

KI in der Arbeitswelt

Miriam Terner
25.06.2026

karriere.at

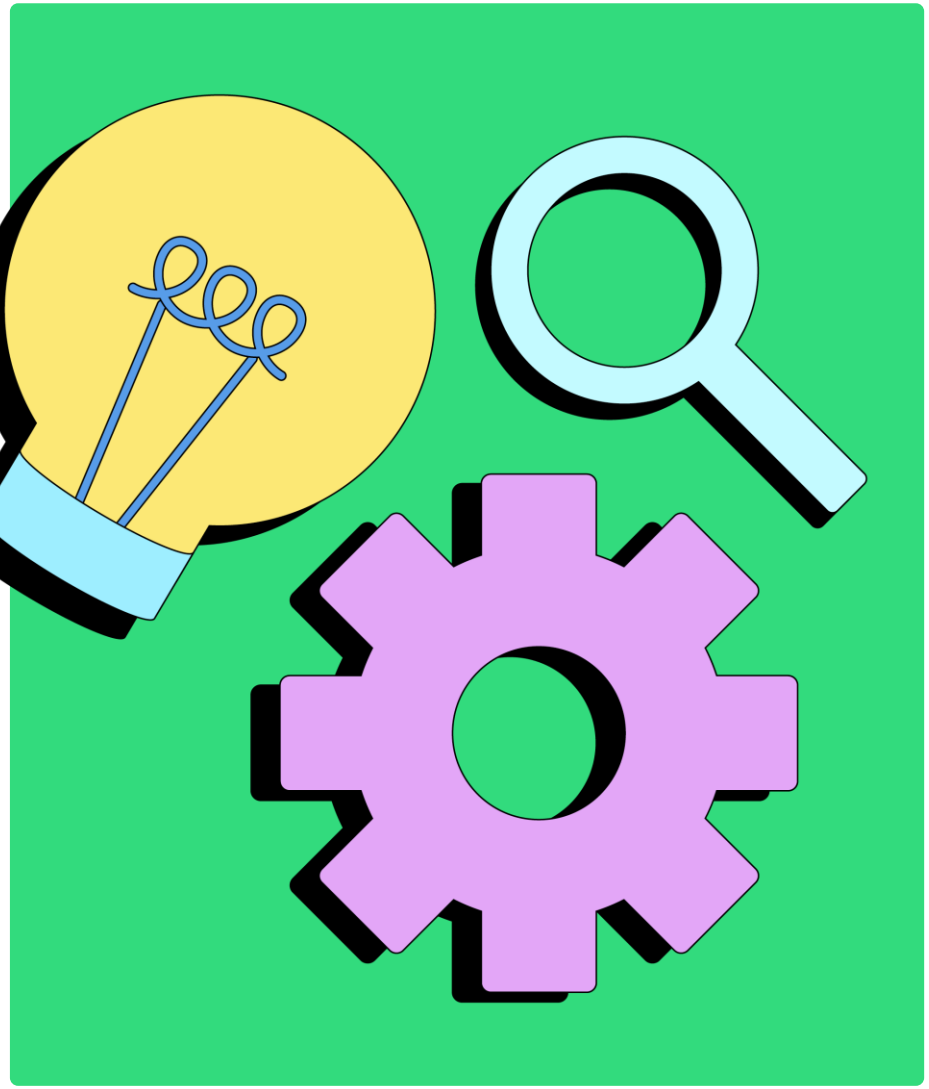
Wer hat diese Woche bereits KI genutzt?



karriere.at

Nicht: Nutzen wir KI?

Stattdessen: Wie nutzen wir sie verantwortungsvoll?



Überraschungen

1

Wartezeit kaum relevant

2

Hohe Zufriedenheit

3

Große Neugier auf Kundenseite

4

Herausforderung: Error Rate

Wenn Menschen betroffen sind, geht es um Fairness, Transparenz, Vertrauen und Verantwortung.

karriere.at

1. Garbage in.

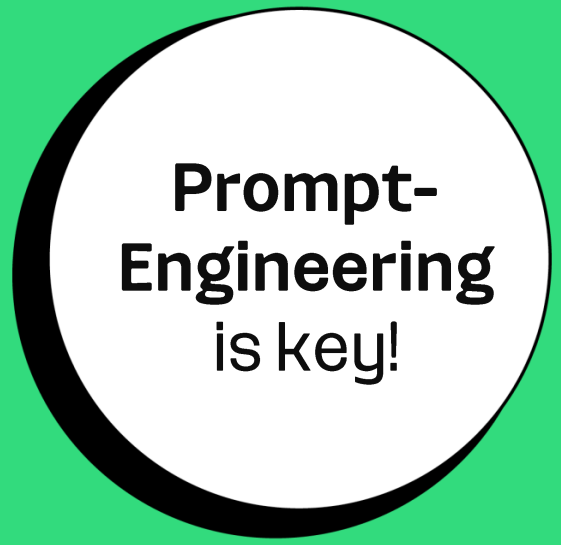
Garbage out.

2. Qualität

Die Qualität des Outputs hängt von der Qualität des Inputs ab.

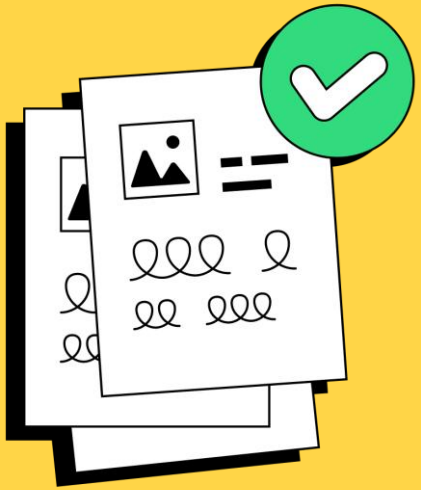
3. Mehr Kontext

Heißt bessere Ergebnisse.

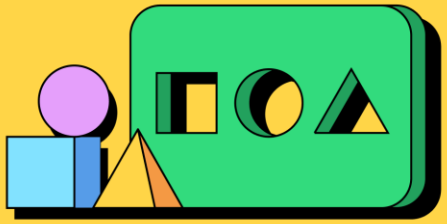


karriere.at

Nähkästchen

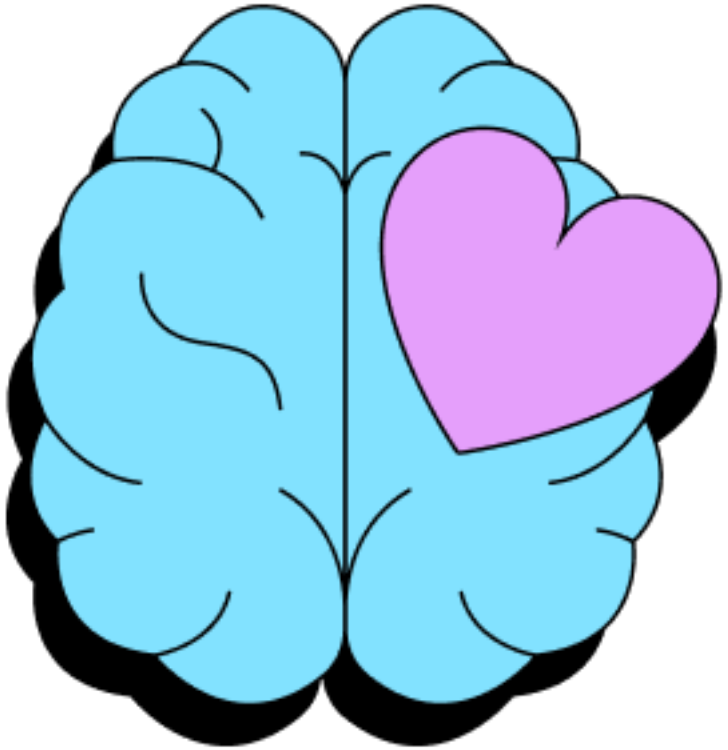


Job Ad Creation



Job Fit Score

karriere.at



Die eigentliche Herausforderung

1. 54 % Match klingt objektiv.
2. Ist es nicht automatisch.
3. Jeder Score basiert auf Annahmen und Gewichtungen.

Wo entsteht Bias?

Mensch ⟷ KI

Der Ursprung ist unterschiedlich.

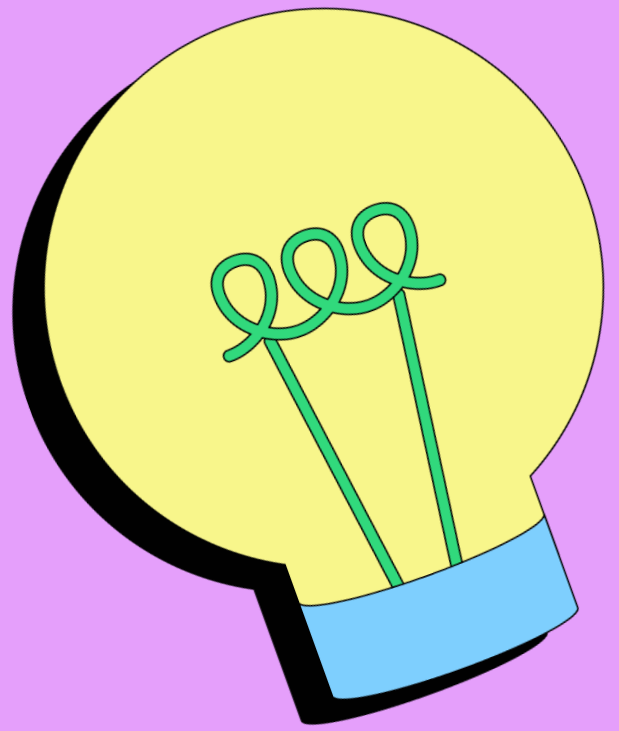
Die Verantwortung bleibt.

Learning

KI macht Entscheidungen nicht automatisch besser.

KI unterstützt menschliche Entscheidungen.

Würdest du ein Talent mit einem Matching Score von 54% zu einem persönlichen Gespräch einladen?



Danke fürs Zuhören!

karriere.at

Good Practice: KNAPP

Markus Otruba
Legal Compliance Management

AI bei KNAPP

Zwischen Effizienz und Verantwortung (oder warum AI mehr braucht als gute Prompts)

Mag. Markus Otruba, MA

Compliance & Datenschutz

KNAPP AG

markus.otruba@knapp.com

www.linkedin.com/in/markus-otruba/

KI unterstützt – ersetzt aber keine Verantwortung

- KI ist bereits Teil des Arbeitsalltags und bietet enorme Chancen sowie Effizienzsteigerungen.
- KI verursacht auch neue Risiken, etwa falsche oder durch Biases verzerrte Outputs, Datenschutzverletzungen, Informationsabfluss, unklare Verantwortlichkeiten.
- Interne (AI-)Policy als Leitfaden für sichere ethische und rechtskonforme Nutzung.
- Das klare Ziel: nicht verhindern, sondern verantwortungsvoll ermöglichen.

Übersicht: AI Policy

- Globale Gültigkeit – auch in Regionen ohne entsprechende Regulatorik.
- Anwendbar für genutzte, entwickelte, implementierte oder vertriebene KI-Systeme.
- Regelt die interne Nutzung sowie die Zusammenarbeit mit Geschäftspartnern und Stakeholdern (in Bezug auf AI).
- Freigabeprozess durch interne Instanz (Perspektive Anbieter oder Betreiber).
- Interne Accountability.
- Nutzungsgebote und –verbote (zB Schatten-KI).

Policy-Inhalte 1/2

- Verboten: inakzeptable Praktikanten gem. Art 5 KI-VO.
- Mit Vorbehalt: Hochrisiko KI-Systeme (grds nicht gestattet).
- Besonders sensibel: HR, Beschäftigung, Bildung
  - Einstellung/Auswahl v. Bewerber:innen
  - Sichtung oder Filterung von Bewerbungen
  - Bewertungen von Personen (Bewerber:innen, Personal…)
  - Entscheidungen zu Beförderungen, Kündigungen
  - Beobachtung und Bewertung von Leistung und Verhalten
- Ja, auch Hochrisiko-KI-Systeme können Prozesse effizienter machen.
  - Aber: sie können auch Karrieren, Chancen und Lebensrealitäten von Menschen beeinflussen.
  - Ethische Vertretbarkeit wenn KI derartige Entscheidungen wie auch immer beeinflusst.

Policy-Inhalte 2/2

- Human in the loop / Menschliche Aufsicht
  - AI assisted – not AI decided. Gerade wenn Ausgaben Wirkungen auf Personen haben.
- Fairness und Nichtdiskriminierung
  - Prüfung von Ausgaben auf Bias & Verzerrungen aus Trainingsdaten
- Transparenz und Erklärbarkeit
  - Menschen von Ausgabe betroffen – höhere Gewichtung von Transparenz als Effizienz
- Verantwortung und Accountability
  - Wer KI einsetzt ist verantwortlich.
- Qualität statt AI-Slop
  - KI-generierte Entwürfe sind kein Arbeitsergebnis sondern bestenfalls Rohmaterial.
- Proportionalität/Verhältnismäßigkeit
  - Nicht jeder KI-Einsatz braucht denselben Governance-Aufwand. Je näher der Use Case an Menschen, sensiblen Daten oder Sicherheitsfragen liegt, desto strenger muss darauf geachtet werden.

KI-Kompetenz – die wichtigste Maßnahme

- KI-Ergebnisse sind nicht automatisch objektiv oder neutral.
- Ausgaben werden durch Trainingsdaten, Prompting, Kontext, Sprache oder unvollständige Informationen verzerrt.
- Mitarbeiter:innen müssen sensibilisiert werden, typische Biases zu erkennen und Ergebnisse nicht ungeprüft zu übernehmen.
- KI-Kompetenz ist nie abgeschlossen – sie muss laufend berücksichtigt werden.
- Viele Nutzer:innen behandeln Ausgaben als fertiges Ergebnis statt als überprüfungsbedürftigen Entwurf.
- KI-Kompetenz bedeutet nicht nur, KI bedienen zu können, sondern KI-Ausgaben kritisch einordnen zu können.

Abschluss

- **KI kann unterstützen, strukturieren und beschleunigen – aber sie kann weder Verantwortung tragen noch ethische Abwägungen für uns übernehmen.**

Ausblick: Termine 2026

13. Oktober 2026 – Von Pride zu Praxis: Allyship

21. Oktober 2026 – re:think diversity. Der Kongress

22. Oktober 2026 – re:think diversity. Die Workshops

Anmeldung auf www.diversitythinktank.at/events & www.rethink-diversity.eu

Vielen Dank unseren Partner:innen!

Vielen Dank fürs Teilnehmen!

Diversity Think Tank Consulting GmbH
Stolberggasse 34
1050 Wien
www.diversitythinktank.at
hello@diversitythinktank.at
www.diversitycampus.eu