



**Code of Conduct  
Demokratische KI**

**White Paper**

# **Solidarische Praxis entlang der Nutzung von KI verankern**



## White Paper

# Solidarische Praxis entlang der Nutzung von KI verankern

## Inhalt

---

1	Handlungsräume eröffnen: KI und Solidarität in der Zivilgesellschaft	3
2	Solidarität in der Praxis: Nutzungszyklus von KI	7
3.	Gemeinsame Standards für mehr Orientierung: Code of Conduct Demokratische KI	29
4.	Impressum	32
5.	Reflexionsvorlagen	33
	Reflexionsvorlage für die Einführung von KI-Tools	34
	Reflexionsvorlage für die Nutzung von KI-Tools	35

# 01 Handlungsräume eröffnen: KI und Solidarität in der Zivil- gesellschaft

Solidarität ist ein zentraler Grundwert gesellschaftlichen Handelns. Im Kontext von KI scheint Solidarität jedoch vor allem ein überfordernd hoher Anspruch zu sein: Wenige globale Unternehmen dominieren die Entwicklung großer KI-Modelle und der digitalen Infrastruktur. Dadurch sind digitale Räume von ungleichen Machtverhältnissen geprägt. Für Organisationen der Zivilgesellschaft entsteht so ein Dilemma zwischen eigenen Werten und realistischen Handlungsräumen.

Dennoch kann die Zivilgesellschaft ihre Rolle als kritische Anwenderin und Gestalterin nutzen, um den Einsatz von KI solidarischer zu gestalten. Durch bewusste Entscheidungen lassen sich konkrete Handlungsräume öffnen: in der Auswahl von KI-Anwendungen und Anwendungsbereichen, im Aufbau von Kompetenzen sowie in der Arbeit an gemeinsamen Standards. Indem Organisationen ihre Lernprozesse und Erfahrungen teilen und eine produktive Fehlerkultur fördern, können sie das Bewusstsein für Probleme wie Verzerrungen und Diskriminierungen schärfen und kritische Stimmen befähigen. Das stärkt die Handlungsfähigkeit anderer Akteur:innen

Warum ist solidarisches Handeln im Kontext von KI überhaupt wichtig? Weil der Einsatz von KI-Systemen grundlegende Herausforderungen mit sich bringt:

- **Unsichtbare digitale Arbeit:** Menschen leisten oft unsichtbare Arbeit, wie Data Labeling oder Annotationen, die KI-Systeme erst ermöglichen.
- **Ökologische Kosten:** Der immense Ressourcenbedarf von KI betrifft oft bereits benachteiligte Regionen, ob durch den Energieverbrauch, die Nutzung von sauberem Wasser oder den Abbau seltener Rohstoffen

## Was bietet das 'White Paper'?

- Eine Analyse der Spannungsfelder zwischen Solidarität und KI.
- Methoden und Reflexionsfragen, um Impulse für eine solidarische KI-Praxis zu entwickeln.
- Eine kuratierte Sammlung von Ressourcen zur vertiefenden Auseinandersetzung.



- **Fehlender Datenschutz und Teilhabe:** Die Kontrolle über Daten wird von großen Firmen dominiert, während KI-gestützte Entscheidungen uns alle betreffen. Besonders vulnerable Gruppen sind dabei häufig stärker von den Auswirkungen betroffen und zugleich nur selten an Entscheidungen über den Einsatz von KI beteiligt.
- **Effizienz narrative statt Gemeinwohl:** (Vermeintliche) Effizienzsteigerungen können die Arbeitswelt verändern und Menschen mit geringen KI-Kompetenzen verdrängen, oft ohne gesellschaftliche Mehrwerte zu schaffen.
- **Ausbeutung kreativer Arbeit:** Journalist:innen, Autor:innen oder Künstler:innen, deren Inhalte für das Training von KI-Modellen genutzt werden, erhalten weder eine angemessene Vergütung noch Anerkennung für ihre Arbeit.

Das Ziel des White Papers ist, Räume für solidarisches Handeln aufzuzeigen und zu öffnen. Dafür beleuchten wir den Entwicklungszyklus von KI, von der Entwicklung und Auswahl über die Nutzung bis zur Evaluation und stellen Fragen und Reflexionsmethoden vor. Die hier formulierten Überlegungen sind ein Anfang, um konkretere Ansätze für solidarisches Handeln zu identifizieren. Auch wenn sich die Spannungen nicht auflösen lassen, können kleine Inseln von Solidarität Möglichkeiten zur Mitgestaltung schaffen.

### Was bedeutet Solidarität?

Solidarität heißt, gemeinsam Verantwortung zu tragen und andere überlegt und wirksam zu begleiten. Es bedeutet, die Interessen und Lebensrealitäten aller zu berücksichtigen. Insbesondere dort, wo Stimmen und Perspektiven strukturell weniger sichtbar sind, braucht es Aufmerksamkeit und Raum für deren Ausdruck.

Solidarität erfordert ein Verständnis der Auswirkungen auf andere und setzt das Wahrnehmen von Verantwortlichkeiten und Zusammenhängen voraus. Solidarität ist somit eine bewusste Praxis, ein Zugehörigkeitsgefühl gegenüber anderen Mitgliedern einer Gesellschaft, das darauf ausgerichtet ist, Bedingungen zu schaffen, in denen vielfältige Bedarfe und Bedürfnisse selbstbestimmt eingebracht werden können.<sup>1</sup>

Während andere Werte wie Gerechtigkeit, Inklusivität oder Verantwortung über rechtliche Rahmen und institutionelle Mechanismen (z. B. durch den EU AI Act) zumindest teilweise herstellbar sind, wird Solidarität in internationalen KI-Richtlinien selten explizit benannt.<sup>2</sup> Eine Ausnahme bildet die Erklärung von Montreal für eine verantwortungsvolle Entwicklung von Künstlicher Intelligenz.

---

<sup>1</sup> Rudschies, C. (2023). Exploring the concept of solidarity in the context of AI: An ethics in design approach. Digital Society, 2(1), S. 1-18. <https://doi.org/10.1007/s44206-022-00027-x>

<sup>2</sup> Jobin, Anna, Marcello Lenca, and Effy Vayena. (2019). „The global landscape of AI ethics guidelines.“ Nature machine intelligence 1.9 S. 389-399 <https://doi.org/10.1038/s42256-019-0088-2>

Sie betont: „Die Entwicklung von KI-Systemen muss in einer Weise erfolgen, dass die Solidaritätsbande zwischen Menschen und zwischen Generationen aufrechterhalten werden.“<sup>3</sup> Solidarität im Kontext von KI bedeutet daher nicht, dass KI-Systeme Empathie entwickeln, sondern dass Menschen in ihrer Gestaltung, Auswahl und Nutzung Solidarität als Prinzip verankern, ob bei der Verteilung von Gewinnen oder bei der Berücksichtigung entstehender Ungleichheiten.

### **Spannungsfeld: Gemeinwohlorientierte Organisationen und Solidarität im Umgang mit KI**

Gemeinwohlorientierte Organisationen bewegen sich im Umgang mit KI in einem besonderen Spannungsfeld: Da Daten, Algorithmen, Rechenressourcen und Infrastrukturen durch wenige Unternehmen kontrolliert werden, sind Organisationen oft gezwungen, kommerzielle Systeme zu nutzen. Gleichzeitig haben sie Ansprüche, solidarisches Handeln konsequent mitzudenken und vorzuleben.

Gemeinwohlorientierte Organisationen stehen daher häufig vor zwei Alternativen: Entweder sie nutzen Tools, deren Entstehungsbedingungen sie nicht kontrollieren können (z.B. schlechte Arbeitsbedingungen von Clickworker:innen, Energieverbrauch, Monopolstrukturen und Diskriminierungsrisiken) oder sie verzichten auf eine Nutzung und sind mit möglichen Konsequenzen, wie erschwerten Aufbau von praktischen KI-Kompetenzen oder gegebenenfalls geringerer Attraktivität für Fördergeber:innen, konfrontiert.

---

<sup>3</sup> Montréal Declaration for a Responsible Development of Artificial Intelligence. (2018). Erklärung von Montréal für eine verantwortungsvolle Entwicklung von Künstlicher Intelligenz [PDF]. S.11 [https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/AL-UdeM\\_Declaration-IA-Resp\\_allemand\\_vf.pdf](https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/AL-UdeM_Declaration-IA-Resp_allemand_vf.pdf)



## Wie kann man solidarische Praxis fördern?

Solidarisch handeln heißt, Ungerechtigkeiten anzuerkennen, auch dort, wo man sie nicht unmittelbar beheben kann, und daraus Konsequenzen für das eigene Handeln abzuleiten. Sie entsteht durch geteilte Verantwortung, gemeinsamen Aufbau von Wissen und die Mobilisierung für Veränderung

- **Nicht-Nutzung als Option ernsthaft beleuchten und bewusste Entscheidungen treffen:** Solidarisch handeln heißt auch, bestimmte Tools nicht einzusetzen, wenn die Kosten für andere (Menschen, Umwelt, demokratische Strukturen) zu hoch sind. Eine bewusste Abwägung mit ehrlicher Berücksichtigung der Kosten ist besser als wegzuschauen. Diese Solidarität umfasst auch, gegen den Zwang zur Nutzung von KI-Systemen einzustehen.
- **Anerkennung von unsichtbarer Arbeit und Ressourcenverbrauch durch politische Arbeit und gemeinsame Forderungen:** Wir tragen die Verantwortung, faire Bedingungen entlang der digitalen Wertschöpfungskette einzufordern und Arbeit wie Data Labeling transparent zu machen.
- **Reflexion und Teilhabe:** Räume schaffen, in denen Mitarbeitende, Ehrenamtliche und Betroffene über Nutzen und Risiken mitentscheiden. Dies beinhaltet die transparente Kommunikation eigener Herausforderungen, um gemeinsame Lernprozesse zu ermöglichen.
- **Beziehungen bewusst gestalten:** Der Einsatz von KI soll nicht auf die Ersetzbarkeit menschlicher Arbeit abzielen, sondern Freiräume für Beziehungsarbeit, Fürsorge oder strategische Tätigkeiten ermöglichen und so Effizienzgewinne sinnvoll nutzen.
- **Zusammenarbeit stärken und Ressourcen bündeln:** Durch gegenseitige Hilfe und Unterstützungsangebote können Gewinne von KI, ob ökonomische oder zeitliche, gleichwertiger verteilt werden. Durch die Entwicklung gemeinsamer Standards können Organisationen Orientierung und Handlungsfähigkeit für Mitarbeitende und Aktive und Glaubwürdigkeit gegenüber externen Organisationen und Fördergeber:innen schaffen.

## 02 Solidarität in der Praxis: Nutzungszyklus von KI

Solidarische Praxis beginnt bei konkreten Entscheidungen, die prägen, wie technische Systeme für und mit Menschen genutzt werden: Wer wird gehört? Wessen Bedürfnisse zählen? Wer trägt welche Verantwortung? Das Verständnis von Solidarität als Gestaltungsprinzip von KI-Systemen beruht auf einem soziotechnischen Verständnis von KI: soziale Machtverhältnisse und Normen prägen technologische Systeme und werden gleichzeitig von ihnen geprägt.

Mit diesem White Paper möchten wir daher dazu einladen, die Auswirkungen dieser Entscheidungen abzuwägen. Einerseits mit Reflexionsfragen und Methoden, andererseits mit Hinweisen zu bestehenden Ressourcen. Dabei orientieren wir uns an vier Phasen des Entwicklungs- und Nutzungszyklus von KI. Denn Solidarität sollte entlang der ganzen Wertschöpfungskette von KI mitgedacht werden, von der Entwicklung und Gestaltung über die Auswahl, bis zum Einsatz und der Evaluation.

### **I. Entwicklung: Wie werden KI-Anwendungen entwickelt und für wen?**

In der Entwicklung von KI-Systemen ist die Machtkonzentration besonders ausgeprägt. Nur wenige Organisationen der Zivilgesellschaft sind direkt an der Entwicklung von Modellen beteiligt oder können diese mitgestalten. Daher tragen besonders die Organisationen, die ausreichende Expertise und Ressourcen haben, eigene Systeme zu entwickeln, eine besondere Verantwortung, die gesamte Wertschöpfungskette von KI in den Blick zu nehmen: von den ethischen Prinzipien über die Fairness der Trainingsdaten bis hin zu den Arbeitsbedingungen der Menschen, die diese Daten annotieren.

Unabhängig vom Umfang der Beteiligung, ob bei eigenen Anpassungen wie Finetuning oder durch Forderungen an Anbieter:innen, bedeutet Solidarität in der Entwicklungsphase, bestehende Machtverhältnisse zu hinterfragen, die Bedürfnisse marginalisierter Gruppen mitzudenken und KI-Systeme so zu gestalten, dass sie einen gesellschaftlichen Nutzen hervorbringen. Auch für Organisationen, die keine eigenen Anwendungen entwickeln, ist es sinnvoll, das Bewusstsein für diese Herausforderungen zu stärken und Menschenzentrierung, Teilhabe und Transparenz einzufordern.



## Lösungsansätze

- **Bias-Reduktion durch diverse Trainingsdaten:** Eine Datenbasis, die die Diversität der Zielgruppe realistisch abbildet, ist notwendig, um die Reproduktion von Stereotypen und Diskriminierungen zu verringern. Dabei ist Vielfalt eine notwendige Voraussetzung für zuverlässige Ergebnisse, allerdings keine Garantie.
- **Datenbestände teilen:** Durch die Aufbereitung und sichere Veröffentlichung von Datenbeständen sowie die gemeinsame Datennutzung werden hochwertige und vielfältige Datensätze für unterschiedliche Organisationen zugänglich und nutzbar. So können besser an die Zivilgesellschaft angepasste und gemeinwohlorientierte KI-Systeme gefördert werden.
- **Transparenz:** Durch Transparenz über Datensets und Algorithmen werden Verzerrungen und Interaktionen verschiedener Merkmale überprüf- und korrigierbar. So lässt sich auch die Gemeinwohlorientierung stärken.
- **Einbindung marginalisierter Perspektiven:** Um blinde Flecken zu vermeiden und die praktische Relevanz zu gewährleisten, ist die direkte Beteiligung von Zielgruppenvertreter:innen notwendig. Aktive Beteiligung, zum Beispiel durch Co-Design-Prozesse oder Fokusgruppen, ist der Simulation von Perspektiven (z. B. über Diversity Personas)<sup>4</sup> vorzuziehen.
- **Solidarisch mit unsichtbarer Arbeit:** Viele KI-Systeme beruhen auf menschlicher Arbeit, wie Content Moderation, Annotation oder Datenmanagement. Clickworker:innen führen diese Arbeit oft unter ausbeuterischen Bedingungen aus. Daher ist es umso wichtiger, bei eigener Entwicklung auf die Entstehung der Trainingsdaten zu achten und bessere Bedingungen einzufordern.

---

<sup>4</sup> KI-Campus. (o.J.). Methoden der Bias-Reduktion für eine sozialverantwortliche KI-Gestaltung: Abgerufen am 8. Dezember 2025, von <https://ki-campus.org/kurse/ki-bias>





## Ressourcen und Werkzeuge: Entwicklung

Kategorie	Ressource	Beschreibung	Link
Rahmenwerke	Algo Rules	9 Regeln zur Gestaltung algorithmischer Systeme mit Fokus auf Transparenz, Verantwortung und Nachvollziehbarkeit.	<a href="#">Zur Ressource</a>
	Critical Engineering Manifesto	Rahmenwerke, die zur Reflexion über Technologie und gesellschaftliche Machtverhältnisse in der Entwicklung auffordern, Fokus auf Verantwortung von Entwickler:innen.	<a href="#">Critical Engineers</a>
	Holberton-Turing Oath		<a href="#">Holberton-Turing Oath</a>
	IEEE Ethically Aligned Design		<a href="#">IEEE Ethically Aligned Design</a>
Selbstbewertung	AI Ethics Impact Group (Ethik-Label A-G)	Entwicklung eines Ethik-Labels mit Unterkategorien wie Privatsphäre oder Fairness. Hilfreich auch zur Bewertung von Dritt-anbieter-Tools.	<a href="#">Zum Bericht</a>
	Assessment List for Trustworthy Artificial Intelligence (ALTAI)	Werkzeug der High Level Expert Group zur Selbstbewertung, u. a. in Bezug auf Transparenz, Governance und gesellschaftliche Auswirkungen.	<a href="#">Zur Liste und dem Tool (engl.)</a>
	OECD (Überblick Tools und Metriken)	Bietet einen Überblick verschiedener Tools und Metriken zur Unterstützung verantwortungsvoller KI.	<a href="#">Zum Überblick</a>



Kategorie	Ressource	Beschreibung	Link
<b>Planung</b>	Planung von Datenvorhaben (Civic Data Lab)	Miro Board, unterstützt bei der strukturierten Planung von größeren Vorhaben mit Daten	<a href="#">Zum Miroboard</a>
	Tarot Cards of Tech	Befassen sich mit Konsequenzen und Möglichkeiten von Entwicklungen und helfen, Lösungen diverser und mit mehr Impact zu gestalten.	<a href="#">Zur Website (eng.)</a>
<b>Rechtliche Orientierung</b>	EU AI Act Compliance Journey (appliedAI Institute for Europe gGmbH)	Unterstützt Organisationen dabei, rechtliche Anforderungen der EU-KI-Verordnung zu verstehen.	<a href="#">Zur Compliance Journey (eng.)</a>
	EU AI Act Compliance Checker (Future of Life Institute)	Gibt rechtliche Orientierung entlang der EU-KI-Verordnung.	<a href="#">Zum Compliance Checker</a>
<b>Daten-Governance &amp; Transparenz</b>	Vorlage des europäischen AI Office	Vorlage zur Offenlegung der Trainingsdaten von Basismodellen.	<a href="#">Zur Vorlage</a>



## Reflexionsfragen

- Wer definiert das Problem, das wir mit KI lösen wollen und wessen Perspektive fehlt?
- Wie reflektieren wir Machtverhältnisse in der Team- und Projektstruktur?
- Beziehen wir systematisch Menschen ein, die von der Lösung betroffen sein könnten, insbesondere benachteiligte Gruppen?
- Berücksichtigen wir in der Datenbeschaffung und -auswahl mögliche strukturelle Verzerrungen oder Ausschlüsse?
- Wie wird unsichtbare Arbeit wie Annotation, Datensäuberung oder Testing sichtbar gemacht?
- Wie gehen wir mit Inhalten oder Daten um, von denen wir nicht wissen, woher sie stammen?
- Kann das Projekt Ungleichheiten reduzieren oder könnte es diese ungewollt verstärken?
- Haben wir uns bewusst mit anderen Lösungswegen auseinandergesetzt und zum Beispiel Ressourcen berücksichtigt?
- Würden wir diese Anwendung auch als „solidarisch“ empfinden, wenn wir selbst zu den Betroffenen gehörten?

## II. Auswahl: Welche Anwendung wird ausgewählt und wie läuft das ab?

Die reflektierte Auswahl konkreter KI-Anwendungen und Anwendungsfälle ist ein zentraler Hebel, um Solidarität in die Praxis zu bringen. Denn die Entscheidung für oder gegen ein bestimmtes KI-Tool oder Modell hat Auswirkungen, die über Budget und technische Voraussetzungen hinausgehen. Wer auswählt, gestaltet mit: Wenn Organisationen KI-Anwendungen übernehmen, ohne zu wissen, wie sie entwickelt wurden oder welche Risiken sie mit sich bringen, riskieren sie, technologische Abhängigkeiten zu verstärken, Diskriminierungen oder Stereotype zu reproduzieren oder zur Verbreitung von Fehlinformationen beizutragen. Durch die Festlegung klarer Ausschlusskriterien für sensible Bereiche können eigene Werte gewahrt werden.

## Lösungsansätze

- **Nutzung offener Modelle oder Angebote gemeinwohlorientierter Organisationen:** Durch die Nutzung von offenen (Open Source) oder teiloffenen (Open Weight) Modellen wird die Überprüfbarkeit von KI-Systemen erleichtert. Open-Weight-Modelle enthalten zwar Informationen über die trainierten Parameter und lassen sich relativ leicht anpassen, jedoch liegen im Gegensatz zu Open-Source-Modellen meist weder die ursprünglichen Trainingsdaten noch der vollständige Code vor. Trotzdem bieten sie mehr Transparenz als vollständig geschlossene Systeme. Die Anpassbarkeit von Open-Source-Modellen kann außerdem technologische Abhängigkeiten von globalen Unternehmen reduzieren.
- **Prüfung von Standards bei der Entwicklung:** Organisationen können prüfen, welche Standards Anbieter:innen bereits in der Entwicklung berücksichtigt haben, zum Beispiel durch den General-Purpose AI Code of Practice der EU.<sup>5</sup> So können sie die Wahrscheinlichkeit verringern, Tools mit inhärenten Diskriminierungen zu implementieren. Zertifizierungen für vertrauenswürdige KI-Systeme, die Nutzende bei der Auswahl unterstützen könnten, stehen allerdings noch am Anfang.<sup>6</sup>
- **Kommunikation der Entscheidungskriterien:** Transparenz über den Auswahlprozess sowohl innerhalb der Organisation als auch außerhalb erfordert einerseits eine gründliche Abwägung und erleichtert andererseits auch externen Organisationen die Entscheidung.
- **Auswahl der (Nicht-)Anwendungsbereiche:** Durch das gemeinsame Festlegen von klaren Grenzen können Organisationen Risiken verringern, etwa in sensiblen Bereichen wie Recruiting oder bei ressourcenintensiven Anwendungen wie Bildgeneration. Dies erfordert die bewusste Auswahl von Anwendungsbereichen, in denen KI eingesetzt beziehungsweise nicht eingesetzt wird, sowie eine problemzentrierte Herangehensweise.

---

<sup>5</sup> Für weitere Standards siehe auch Phase Entwicklung.

<sup>6</sup> Baeva, G., Brunner, R., Elmas, F., Fresz, B., Fuß, C., Heß, C., ... & Seyerlein-Klug, A. (2023). Welche Zertifizierung ist hilfreich für Verbraucher\* innen?. <https://doi.org/10.24406/publica-2132>

## Solidarität als bewusste Entscheidung: Warum die Digitale Gesellschaft e.V. auf die Nutzung generativer „KI“ verzichtet

Konstantin Macher, Digitale Gesellschaft e. V.

Als Digitale Gesellschaft e.V. beschäftigen wir uns schon seit 15 Jahren damit, wie Digitalpolitik, Grundrechte und demokratische Teilhabe zusammenhängen und engagieren uns dafür, diese im Sinne von Menschen zu gestalten. Ob wir wollen oder nicht, das Thema „KI“ hat eine gesellschaftliche Relevanz erreicht, die eine (kritische) Auseinandersetzung damit erfordert.

Wir haben ausführlich abgewogen, ob wir den Code of Conduct Demokratische KI (Anm. Einführung in Kapitel 03) mitzeichnen und die Prinzipien des Code of Conducts für uns anwenden können. Die Schlussfolgerung: JA, wir haben mitgezeichnet und empfinden den Code of Conduct als ein wertvolles Instrument, um diese Auseinandersetzung in der Organisation anzugehen. Und NEIN, generative „KI“ wollen wir in unserer Arbeit auf absehbare Zeit nicht nutzen. Anhand drei der verankerten Prinzipien im Code of Conduct und anhand unseres Verständnisses von Solidarität möchte ich kurz erklären, wie wir zu dieser Schlussfolgerung gekommen sind:

**Abwägung der Nutzung:** „KI“ ist für uns kein Selbstzweck. Wir wollen uns nicht vom Hype treiben lassen. Ganz im Gegenteil: Die renommierten Expertinnen bei der Forschung zu „KI“, die Linguistin Emily Bender und die Informatikerin Timnit Gebru haben „KI“-Sprachmodelle den Ausdruck „stochastische Papageien“<sup>7</sup> geprägt: „KI“ ist hier das Nachplappern abstrakter Daten ohne Sinn und Verstand. Gleichzeitig gehen damit große potenzielle Schäden einher. Als Digitale Gesellschaft wägen wir darum ab, ob der Einsatz von „KI“-Anwendungen für unser Ziele sinnvoll ist, welche negativen Konsequenzen für Menschen und Umwelt damit möglicherweise verbunden sind und ob uns dann der Einsatz in der Abwägung als vertretbar erscheint.

**Menschenzentrierung:** Wir stellen die Erfahrungen, Bedürfnisse und Interessen von Menschen in den Vordergrund und wie sich die Entwicklung und der Einsatz von „KI“ auf diese auswirkt. Dabei berücksichtigen wir Ressourcen, Arbeitsbedingungen und Energieverbrauch entlang der gesamten Lieferkette. „KI“-Systeme basieren auf menschlicher Arbeit, insbesondere der beteiligten „Data Workers“. Diese Arbeit bleibt oft unsichtbar und wird häufig unter hochgradig ausbeuterischen Bedingungen geleistet.

**Diskriminierungskritische Haltung:** Die Entwicklung von „KI“-Modellen basiert auf Trainingsdaten. Die darin abgebildeten gesellschaftlichen Verzerrungen und Diskriminierungen werden bei der Anwendung reproduziert und verstärkt. Wir priorisieren bei unserer Abwägung die Interessen von Menschen, deren strukturelle Benachteiligungen durch den Einsatz von „KI“ noch verstärkt würde.

Wir lieben Technik und experimentieren gerne, aber reflektieren dabei auch unsere Verantwortung. Unsere Abwägung hat anhand der Prinzipien des Code of Conducts ergeben: die möglichen Vorteile rechtfertigen für uns keine Nutzung generativer „KI“ in unserer Arbeit.

Wir möchten mit dieser Entscheidung auch anderen Mut machen: die Nichtnutzung von „KI“ ist auch eine Option.

---

<sup>7</sup> Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).



## Ressourcen und Werkzeuge: Entwicklung

Kategorie	Ressource	Beschreibung	Link
<b>Vergleich &amp; Analyse</b>	Datenbank von Huggingface (Offene Modelle)	Ermöglicht den Vergleich verschiedener offener Modelle. Filter erlauben die Limitierung auf bestimmte Größen oder Anwendergruppen und erleichtern die Überprüfbarkeit der Modelle.	<a href="#">Zur Datenbank</a>
<b>Governance &amp; Standards</b>	AI Safety Index des Future of Life Institute	Bewertet regelmäßig große KI-Firmen, insbesondere in Bezug auf Risikomanagement, Sicherheit, Transparenz und Governance. Bietet eine fundierte Grundlage für Auswahlentscheidungen.	<a href="#">Zum Index</a>
	General-Purpose AI Code of Practice der EU	Eine freiwillige Selbstverpflichtung von Firmen zu bestimmten Mindeststandards von der EU, partizipativ entwickelt.	<a href="#">Zur Website</a>
<b>Auswahl von Anwendungsbereichen</b>	KI-Wertekompass (White Paper KI und Freiheit)	Hilft Organisationen, ihre eigenen Werte beim Einsatz von KI zu reflektieren. Dies unterstützt eine Tool-Auswahl, die den eigenen Wertvorstellungen entspricht.	<a href="#">Zum White Paper (S.7-12.)</a>
	Entscheidungsflow (White Paper KI und Gerechtigkeit)	Hilft Organisationen, die Nutzung von KI-Anwendungen in einem bestimmten Bereich abzuwägen.	<a href="#">Zum White Paper (S. 19)</a>



Kategorie	Ressource	Beschreibung	Link
	Data Heroes (Civic Data Lab)	Spiel zur Vermittlung von Datenkompetenz und kritischem Umgang mit KI	<a href="#">Zum Spiel</a>
	Digitalisierung vs. Digitalisieren (Civic Data Lab)	Blogartikel mit Hinweisen zur Stärkung einer guten Datenkultur	<a href="#">Zum Blogartikel</a>
	KI-Kompass (Alexander von Humboldt Institut für Internet und Gesellschaft)	Ein Tool, das einen Einstieg und Druckdateien bietet, um zu diskutieren, welche Herausforderungen durch den Einsatz von KI besser bewältigt werden können.	<a href="#">Zum Spiel</a>

### Reflexionsfragen

- Welches Problem löst das Tool?
- Wer hat das Tool entwickelt und wer verdient daran?
- Gibt es Hinweise auf faire Bedingungen in der Entstehung (z. B. Vergütung von Clickworker:innen, Umgang mit Trainingsdaten)?
- Haben wir geprüft, ob es Alternativen gibt, die Open Source oder gemeinwohlorientiert funktionieren?
- Ist das Tool zugänglich für verschiedene Zielgruppen?
- Was wäre das Schlimmste, was die KI-Anwendung anrichten könnte? Was wäre das beste Ergebnis?



- Welche Chancen eröffnet der mögliche KI-Einsatz, unsere Prozesse kritisch zu prüfen und gegebenenfalls zu vereinfachen oder zu verbessern?
- Können wir durch unsere Auswahl andere Gruppen oder Organisationen in ihrem Zugang unterstützen?
- Erhöhen wir durch die Auswahl technologische Abhängigkeiten?
- Gibt es solidarische Bezahl- oder Lizenzmodelle? Wenn ja, unterstützen wir diese?
- Ist das Tool anpassbar für unterschiedliche soziale Kontexte?
- Kommunizieren wir unsere Auswahlentscheidung transparent innerhalb und außerhalb der Organisation?
- Teilen wir unsere Erfahrungen mit anderen, z. B. über Netzwerke oder durch Veröffentlichungen?

### III. Einsatz: Was macht KI in meiner Organisation und darüber hinaus?

Erst beim Einsatz von KI entscheidet sich, wer im Alltag wie von bestimmten Systemen profitieren kann. Im zivilgesellschaftlichen Kontext wird KI vor allem mit dem Ziel eingeführt, zur Entlastung von Prozessen beizutragen, etwa um Anfragen schneller zu beantworten oder Ehrenamtliche zu koordinieren.

Solidarität in dieser Phase bedeutet, in drei Handlungsfeldern aktiv zu werden. Erstens, Mitarbeitende und Aktive im Umgang mit KI zu befähigen. Zweitens, eine kontinuierliche Reflexion über Verantwortlichkeiten bei der Nutzung sicherzustellen. Drittens, dafür zu sorgen, dass entstehende Effizienzgewinne für mehr Kreativarbeit und Beziehungsarbeit genutzt werden können.

#### Lösungsansätze

- **Mitarbeiter:innenbeteiligung:** Beteiligungsformate wie Workshops oder regelmäßige Feedbackrunden ermöglichen es Mitarbeitenden, frühzeitig in die Auswahl, Einführung und Bewertung von KI-Systemen eingebunden zu werden.



### Vulnerabilität reflektieren: Risiko-Matrix

Bei der Einführung von Technologien können besonders für vulnerable Gruppen unbeabsichtigte Risiken entstehen, unabhängig davon, ob sie die Systeme selbst nutzen oder von ihren Auswirkungen betroffen sind. Um Organisationen frühzeitig dabei zu unterstützen, solche Risiken zu erkennen, eignet sich die folgende Matrix. Auf der einen Achse steht die Wahrscheinlichkeit, mit der eine bestimmte Auswirkung eintreten könnte, auf der anderen die Schwere der möglichen Auswirkungen auf die betroffene Gruppe. Diese Form des strukturierten Brainstormings unterstützt Organisationen dabei, ihre Zielgruppen systematisch mitzudenken und Solidarität als praktisches Handlungskriterium zu verankern.

## Vulnerabilitäts-Risiko-Matrix

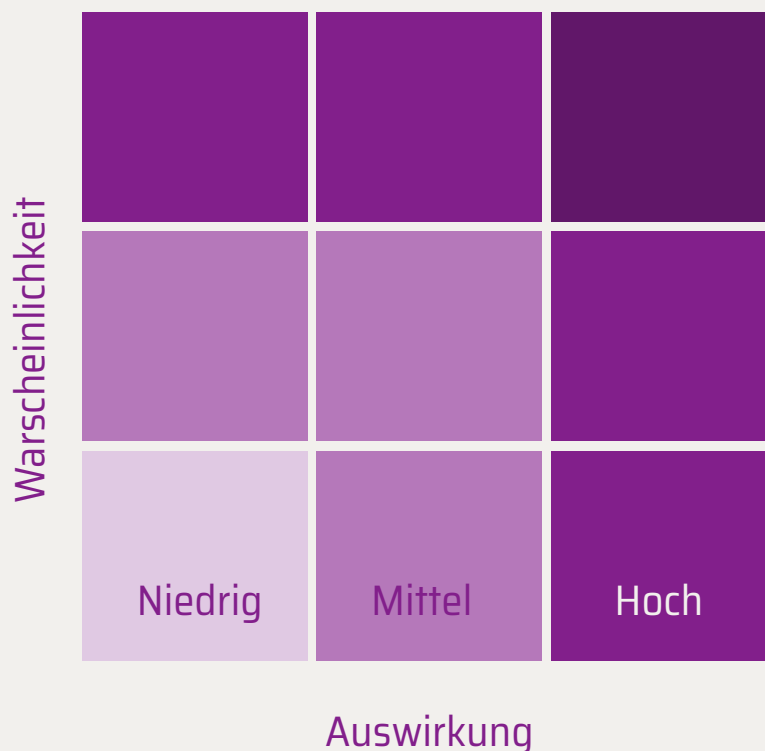


Abb. 1. Vulnerabilitäts-Risiko-Matrix nach Christoph Hassler (H:DEI).



### Workshopkonzept: Toolauswahl in der eigenen Organisation

**Ziel:** Die Auswahl von KI-Anwendungen soll den Werten der Organisation entsprechen und die Bedürfnisse diverser Zielgruppen abbilden. Mögliche Risiken sollen durch Perspektivwechsel minimiert werden.

**Zielgruppe:** Entscheidungsträger:innen, Mitarbeitende, ggf. Aktive oder Zielgruppen

**Dauer:** ca. 2,5 bis 3 Stunden

**Voraussetzung:** Die grundsätzliche Notwendigkeit einer KI-Anwendung für das konkrete Problem ist bereits abgewogen worden.

**Vorbereitung:** Vorab eine engere Auswahl von 2–3 potenziellen Tools treffen.

1. Einstieg (10–15 Min):
  - a. Kurze Vorstellung der Zielsetzung.
  - b. Wie kommt ihr an? Welche Ängste oder Hoffnungen habt ihr?
2. Gemeinsame Visionen entwickeln: Wie passen die KI-Anwendungen zu unseren Organisationswerten? Wo sehen wir rote Linien? Welche Ziele wollen wir erreichen? Welche Methoden wenden wir an, um unsere Ziele zu erreichen? (30–60 Min, ggf Aufteilung in Kleingruppen)
3. Perspektivwechsel (45–60 Min):
  - a. Erstellung realistischer aber diverser Nutzer:innenprofile bzw. Personas (mit Variation bei z.B. Wohnort, Herkunft, Alter, Geschlecht, Digitalisierungskompetenzen, Infrastruktur, Wahrscheinlichkeit der Nutzung)
  - b. In Kleingruppen wählt jede Gruppe eine Persona aus und diskutiert Hürden, Risiken und spezifischen Nutzen.
  - c. Verortung der Persona in der Risiko-Matrix
  - d. Reflexion im Plenum: Welche Perspektive fehlt? Was sind die Auswirkungen auf vulnerable Gruppen und Nicht-Nutzende?
4. Gemeinsam Reflexionsfragen (s.o.) nach den Diskussionen anpassen und diskutieren (30–45 Min)
5. Nächste Schritte definieren (10–30 Min)

## Lösungsansätze

- **AI und Data Literacy:** Ein gemeinsames Verständnis darüber, was KI ist (und was nicht), wie sie funktioniert und welche Grenzen sie hat, ist Voraussetzung für informierte Entscheidungen. Einfache, niedrighschwellige Bildungsangebote stärken das Problembewusstsein und den verantwortungsvollen Umgang.
- **Gemeinsames Lernen:** Durch Erfahrungs- und Wissensaustausch können interne Kompetenzen aufgebaut werden. Das schafft neben Austauschmöglichkeiten innerhalb der Zivilgesellschaft auch mehr Unabhängigkeit von externen Dienstleistern. Beispielformate könnten Stammtische oder kollegiale Fallbesprechungen sein.
- **Solidaritätsgetriebener Einsatz:** Viele Tools können für unterschiedliche Zwecke verwendet werden. Es ist daher entscheidend, den Einsatzbereich bewusst zu kontextualisieren: Organisationen können beispielsweise auf den Einsatz ressourcenintensiver KI-generierter Bilder verzichten und stattdessen dieselbe Technologie nutzen, um interne Wissensbestände durchsuchbar zu machen.

## Solidarität durch gemeinsames Lernen im Projekt „KI für ein gutes Altern“

Alexander Engel, BAGSO – Bundesarbeitsgemeinschaft der Seniorenorganisationen e.V.

Solidarität im Kontext von KI bedeutet nicht nur, Risiken und Exklusionsdynamiken zu reflektieren, sondern auch Chancen für gemeinschaftliches Handeln sichtbar zu machen. Ein Beispiel ist das BMBFSFJ-geförderte BAGSO-Projekt „KI für ein gutes Altern (2023-2025)“<sup>8</sup>:

Von Kiel bis Memmingen, von Bocholt bis Görlitz haben sich seit 2020 bereits 58 Senioren-Internet-Initiativen, Mehrgenerationenhäuser, Seniorenbüros und weitere Einrichtungen, die älteren Menschen digitale Kompetenzen vermitteln, zu „KI-Lernorten“ für ältere Menschen weiterqualifiziert. Die beteiligten Initiativen nahmen an Grundlagen-Webinaren teil und verfügen über Geräte und Anwendungen, mit denen KI im Alltag praktisch erfahrbar wird. Auch zehn bundes- und landesweit aktive Seniorenorganisationen engagieren sich im Rahmen des Projektes für ein gutes Altern mit KI.

Diese Lernorte eröffnen Räume, in denen ältere Menschen KI kennenlernen und im Austausch mit anderen Hemmungen abbauen und Herausforderungen begegnen können. Dabei entsteht ein solidarisches Miteinander: Wer sich mit den Technologien vertraut gemacht hat, gibt sein Wissen als Multiplikator:in weiter und unterstützt andere beim Einstieg. Aus individueller Unsicherheit wird so ein gemeinschaftliches Lern- und Erfahrungsfeld, in dem ältere Menschen sich gegenseitig stärken und auch über die Risiken von KI-Technologien austauschen können.

Die Verbindung von Technikaneignung und solidarischer Praxis zeigt: Erst der konkrete Kontakt mit KI ermöglicht Teilhabe am gesellschaftlichen Diskurs über Chancen und Grenzen der Technologie. Solidarität entsteht hier, indem niemand allein gelassen wird, sondern gemeinsam Wissen, Verantwortung und Handlungsspielräume geteilt werden.

<sup>8</sup> BAGSO (2025) KI-Lernorte: Landkarte: <https://ki-und-alter.de/ki-lernorte/>

## Ressourcen und Werkzeuge: Entwicklung

Kategorie	Ressource	Beschreibung	Link
<b>Selbsteinstufung</b>	DALI Data Literacy für Erwachsene	Rahmenkonzept zur Beschreibung der allgemeinen Datenkompetenz.	<a href="#">Zum Konzept</a>
	Österreichischer Referenzrahmen für digitale Kompetenzen (DigComp 2.3 AT)	Rahmenwerk zur Definition und Entwicklung notwendiger digitaler Kompetenzen (basierend auf EU-Standards).	<a href="#">Zum Rahmen</a>
	KI-Kompetenzprofile (Bertelsmann Stiftung)	Definiert notwendige fachliche und organisationale Kompetenzen für den Umgang mit KI (entwickelt für die Verwaltung).	<a href="#">Zum Bericht</a>
<b>Kompetenzaufbau</b>	Datenlebenszyklus (Civic Data Lab)	Unterstützt beim zielgerichteten Umgang mit Datenbeständen, bietet eine Selbsteinstufung zu Datenkompetenzen, verweist auf weitere Ressourcen.	<a href="#">Zur Website</a>
	Den KI-Drachen zähmen (Julia Junge)	7-tägiger Newsletter zum kompetenten Umgang mit KI im Alltag für NGOs.	<a href="#">Zur Anmeldung</a>
<b>Praxisleitfäden &amp; Fachwissen</b>	Der Paritätische: Textsammlung KI in der sozialen Arbeit	Konkrete Hinweise und Leitfäden für den Einsatz von KI im sozialen Sektor.	<a href="#">Zur Sammlung</a>



Kategorie	Ressource	Beschreibung	Link
	BAGSO: KI und Nachhaltigkeit	Beleuchtet den Zusammenhang zwischen KI, demografischem Wandel und Nachhaltigkeit.	<a href="#">Zum Aufsatz</a>
	Hans Böckler Stiftung: Mitsprachestrukturen	Studie zu Mitsprachestrukturen für Arbeitnehmer:innen bei der Einführung von KI	<a href="#">Zur Studie</a>
<b>Management-Standards &amp; Risikosteuerung</b>	ISO/IEC 42001:2023 (AIMS)(kostenpflichtig)	Internationaler Standard für KI-Managementsysteme zur strukturierten Einführung und Überwachung von KI-Systemen (relevant für größere Organisationen). <sup>9</sup>	<a href="#">Zur Norm (kostenpflichtig, eng.)</a>
	NIST AI Risk Management Framework	Freiwilliger, risikofokussierter Ansatz zur Abschätzung und Reduzierung potenzieller Schäden beim KI-Einsatz.	<a href="#">Zum Framework (eng.)</a>
<b>Anlaufstellen</b>	KI-Service Desk der Bundesnetzagentur	Öffentliche, deutsche Anlaufstelle für Fragen zu KI und der Umsetzung des EU AI Act (Orientierungshilfe).	<a href="#">Zum Service Desk</a>
	AI Act Service Desk der EU	Anlaufstelle der EU-Institutionen für Fragen im Kontext des EU AI Act.	<a href="#">Zum Service Desk (eng.)</a>

<sup>9</sup>Viele Anforderungen entsprechen auch denen der europäischen KI-Verordnung. Problematisch ist jedoch, dass die Prozesse sehr umfassend und damit insbesondere für kleinere Organisationen kaum zu bewerkstelligen sind. Dadurch entsteht ein eigenes Ökosystem kommerzieller Anbieter, die bei der Umsetzung unterstützen.



Kategorie	Ressource	Beschreibung	Link
	Civic Coding	Regelmäßige Veranstaltungen, außerdem Landkarte von gemeinwohlorientierten KI-Projekten	<a href="#">Zur Website</a>

### Reflexionsfragen

- Ersetzen wir mit dem KI-Tool menschliche Fürsorge, statt sie zu ergänzen oder zu entlasten?
- Wie wird menschliche Arbeit durch die Anwendung unterstützt?
- Wer wird durch die Nutzung des Tools entlastet und wer wird (vielleicht unbeabsichtigt) ausgeschlossen?
- Was sind die Effekte auf Personen, die die Anwendung nicht nutzen wollen oder können?
- Wie nutzen wir freiwerdende Ressourcen oder Zeit?
- Kommunizieren wir gegenüber Nutzer:innen, dass KI-Anwendungen verwendet werden? Wenn ja, wie und in welchem Umfang?

## Partizipative KI-Gestaltung im Arbeitsalltag – Beschäftigte und ihr Erfahrungswissen systematisch einbeziehen

Mario Michael Ottaiano, Weizenbaum Institut

Aktuelle Forschung zeigt ein differenziertes Bild in der Frage der Einbindung von Beschäftigten und ihren Interessenvertretungen wie Betriebsrät:innen: Diese werden im Vergleich zu den einzelnen Beschäftigten bei der KI-Einführung oft außen vorgelassen und nur in wenigen Fällen eingebunden.<sup>10</sup>

Die erfolgreiche Einführung von KI-Anwendungen ist voraussetzungsreich. Sinnvoller KI-Einsatz entsteht dort, wo er einerseits die Bedürfnisse der Beschäftigten reflektiert und zugleich an die Erfordernisse der Organisation angepasst ist. Oft bedeutet KI-Einsatz, je nach Art des Systems, dass sich bestehende Arbeitsprozesse teilweise grundlegend verändern. Bisweilen fehlen hierzu jedoch betriebliche Regelungen oder Rahmenbedingungen wie Betriebsvereinbarungen, die den Einsatz und die Anwendung von KI-Systemen im Arbeitskontext regeln und absichern.

Um auf Seiten der Beschäftigten einerseits das nötige Wissen zu fördern, aber auch die Akzeptanz für die Nutzung von KI-Anwendungen zu erhöhen, muss Mitarbeitenden jedoch eine echte Beteiligung angeboten werden, die ihnen tatsächlichen Einfluss auf die Gestaltung ihrer Arbeitsbedingungen und -prozesse einräumt.<sup>11</sup> Indirekte Beteiligungsansätze durch Betriebs- und Personalräte stellen zwar eine solide Grundlage zur Wahrung und Integration von Beschäftigteninteressen bei Technologie- bzw. KI-Einführungs- und späteren Anwendungsprozessen dar.<sup>12</sup> Oft geraten jedoch Ansätze, die rein auf eine stellvertretend ausgerichtete Interessenvertretung durch Betriebs- und Personalräte abzielen, an ihre Grenzen.

Ziel von Betriebs- und Personalräten, aber auch der Organisationsseite sollte es daher sein zusätzlich auf aktive direkte Partizipationsformen zu setzen, um wichtige Teilperspektiven und Handlungswissen von Beschäftigten in die Gestaltung der KI-Systeme zu integrieren.

Entscheidend dafür ist die Sensibilisierung auf Organisationsseite, dass Beschäftigte grundsätzlich darüber mitentscheiden sollten, wie ihre Arbeitsbedingungen auch im technologischen Wandel gestaltet sein sollten. Zudem haben Organisationen einen Nutzen davon, Beschäftigte von Beginn an bei der Einführung von KI zu beteiligen: Denn schließlich verfügen ihre Beschäftigten oft über besondere Fähigkeiten und spezifisches, kontextbezogenes Erfahrungswissen.<sup>13</sup> Dieses gilt es zu mobilisieren, vor allem in der Digitalisierung und insbesondere bei der Entwicklung und Anwendung von KI-Systemen, die für eine Verbesserung der Arbeit und Entlastung von Beschäftigten sorgen können.

---

<sup>10</sup>Krzywdzinski, M. (2024). Zwei Welten der KI in der Arbeitswelt: Wie Management und Betriebsräte die Einführung und Nutzung von KI-Anwendungen gestalten (No. 39). Weizenbaum Discussion Paper. S.12. Abgerufen am 8. Dezember 2025, von <https://www.weizenbaum-library.de/items/aa3e6c0d-8816-4a3a-bd2f-971910e43fb8>

<sup>11</sup>Ottaiano, M. M. (2025). Neue Herausforderungen für die betriebliche Einführung von generativen KI-Systemen: Zur Rolle von Partizipation und Anerkennung. AIS-Studien, 18(2), 143-162. [https://www.arbsoz.de/ais-studien-leser/430-neue-herausforderungen-fuer-die-betriebliche?file=files/downloads/ais-studien/AIS-25\\_2-11\\_Ottaiano.pdf](https://www.arbsoz.de/ais-studien-leser/430-neue-herausforderungen-fuer-die-betriebliche?file=files/downloads/ais-studien/AIS-25_2-11_Ottaiano.pdf)

<sup>12</sup>Kutlu, Y., Seibold, B., Mugler, W., Kirner, E., & Klatt, S. (2023). Betriebliche Digitalisierungsprozesse: Neue Beteiligungsofferten, neue arbeitspolitische Chancen?. Arbeit, 32(1), 51-74. <https://doi.org/10.1515/arbeits-2023-0004>

<sup>13</sup>Ottaiano, M. M., Schneidmesser, L., & Butollo, F. (2025). KI und Erfahrungswissen: Zur Formalisierbarkeit und Delokalisierung von Facharbeit. ARBEIT – Zeitschrift für Arbeitsforschung, Arbeitsgestaltung und Arbeitspolitik, 34(4), 275-299. <https://doi.org/10.1515/arbeits-2025-0015>



## Reflexionsfragen

- Wie gehen wir mit Widerstand oder Ängsten in der Organisation oder Community um?
- Haben wir Ansprechstellen oder Feedbackformate für Menschen, die sich durch das Tool benachteiligt fühlen?
- Entstehen emotionale Beziehungen oder Abhängigkeiten? Falls ja, wie gehen wir damit um?
- Werden Menschen durch diese Anwendung überwacht?
- Welche Arbeitsschritte werden wie angepasst? Wen betrifft das?
- Durch welche Eingaben können wir mehr Beteiligte mitdenken?

## Solidarische Prompts?

Bei der Eingabe in KI-Systeme können wir bewusst soziale Dynamiken und Verantwortung berücksichtigen.<sup>14</sup> Es ist jedoch entscheidend zu beachten, dass Prompting-Strategien die grundlegenden ethischen und technischen Rahmenbedingungen des Modells nicht verändern. Sie verbessern lediglich die Nutzung innerhalb der bestehenden Grenzen:

- **Verringern von Biases:** Durch bewusst inklusive Formulierungen (zum Beispiel Mitarbeitende statt Mitarbeiter) können Biases und Stereotypen bei bereits entwickelten Modellen verringert werden. Außerdem können Biases in Antworten durch gezielte Prompts verringert werden, etwa indem gezielt nach Stereotypen oder Diskriminierungen und ggf. nach Anpassungen gefragt wird.
- **Erfragen von einfacher Sprache:** Die explizite Anforderung von einfacher oder barrierefreier Sprache im Prompt kann komplexe Sachverhalte zugänglicher machen.
- **Kontext der Nutzung angeben:** Die Spezifizierung des Nutzungskontextes (z.B. die Rolle der Organisation, Werte, Sektor) im Prompt berücksichtigt unbeabsichtigte Folgen und kann Verantwortung fördern.

---

<sup>14</sup>Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., ... & Resnik, P. (2024). The prompt report: a systematic survey of prompt engineering techniques. <https://arxiv.org/abs/2406.06608>



#### IV. Evaluation: Wie wirkt KI in meiner Organisation und darüber hinaus?

In der Evaluationsphase wird deutlich, wie Systeme sich auf die eigene Organisation, Zielgruppen und das Umfeld auswirken. Dies erfordert ein umfassendes Verständnis von Wirkung und Verantwortung.

Eine solidarische Evaluation erfordert daher Transparenz und Inklusion. Dabei geht es einerseits darum, welche Zielgruppen erreicht werden und andererseits darum, welche Auswirkungen die Nutzung auf Gruppen hat, die nicht Teil der Entscheidung oder Nutzung waren. Wenn Organisationen diese Erfahrungen nicht nur teilen, sondern reflektieren, ermöglicht dies kollektives Lernen.

#### Lösungsansätze

- **Klare Dokumentation:** Durch niedragschwellige Feedbackformate, Erlebnisberichte und Kurzpräsentationen können Nutzende, Mitarbeitende und Aktive ihre Erfahrungen teilen und reflektieren. Durch die Einbeziehung verschiedener Gruppen können auch Effekte auf Nicht-Nutzende sichtbar werden.
- **Systematische Messung:** Durch klare Zieldefinitionen und Datenerhebungen können Wirkungen systematisch erfasst werden. Hier kann auch der gesellschaftliche Mehrwert kritisch evaluiert werden.
- **Offene Fehlerkultur:** Von einem offenen Umgang mit Fehlern und unbeabsichtigten Folgen profitieren alle. Durch Peer-Evaluation und Austauschformate mit anderen zivilgesellschaftlichen Organisationen können Erfahrungen reflektiert, Best Practices entwickelt werden. Dadurch können Ressourcen gespart werden.



## Methode für Einzelpersonen: Toolreflexion

Ziel: Tool-Effekte reflektieren und dokumentieren  
Dauer: ca. 20-30 Min  
Teilnehmende: 1-2 Personen

Vorlage zur Toolreflexion siehe Anhang.

## Methode für Teams: Retro zur gemeinsamen Reflexion

Methode für Teams: Retro zur gemeinsamen Reflexion

Ziel: Gemeinsames Check-in zur Wirkung von KI-Tools und Teilen von Erkenntnissen

Zielgruppe: Nutzende (Mitarbeitende, Aktive)

Dauer: 30-45 Min, regelmäßig zur und nach der Tool-Einführung

Vorbereitung: ggf. asynchrone Toolreflexion

Vorlage zur Toolreflexion siehe Anhang.

1. Kurze Vorstellung des Tools: Welches Problem soll durch das Tool gelöst werden, wer soll erreicht werden, was sind die Ziele? (5 Min)
2. Gemeinsamer Check-in zu Wirkung (15 Min)
  - a. Was läuft gut? Was überrascht uns?
  - b. Gibt es Spannungen oder unbeabsichtigte Folgen?
  - c. Welche neuen Ideen oder Bedenken sind entstanden?
3. (Optionaler Fokus auf spezifische Risiken) In Kleingruppen reflektieren Teilnehmende zu verschiedenen Kategorien: Wo gibt es Probleme? Wo gibt es Potenziale? (15 Min)
  - a. Transparenz
  - b. Auswirkungen auf Umwelt
  - c. Betroffene Gruppen und Nicht-Nutzende (hier ggf. Risiko-Matrix miteinbeziehen)
  - d. Bias & Diskriminierung
  - e. Zugänglichkeit und Kosten
4. Festlegung nächster Schritte (10 Min)
  - a. Was ändern wir kurzfristig? Was beobachten wir weiter?



## Reflexionsfragen

- Wie messen wir Erfolg?
- Wer bewertet und wer bleibt außen vor?
- Fragen wir nach Rückmeldungen aus beteiligten und nicht beteiligten Gruppen?
- Wie sind Mitarbeitende und Ehrenamtliche eingebunden?
- Wie werden Nicht-Nutzende in den Evaluationsprozess einbezogen?
- Wird Feedback aus marginalisierten Gruppen ernst genommen? Wie wird es integriert?
- Unter welchen Bedingungen würden wir die Anwendung nicht (mehr) einsetzen?
- Reflektieren wir regelmäßig, ob wir unsere ursprünglichen Ziele und Werte noch erreichen?
- Was sind die Konsequenzen, wenn durch unser KI-System Schaden entsteht? Für wen?
- Erkennen wir in der Evaluation unbeabsichtigte Ausschlüsse?
- Nutzen wir Evaluation als kollektiven Lernprozess oder nur zur Legitimation bestehen der Entscheidungen?
- Teilen wir unsere Erfahrungen offen mit anderen, auch wenn sie problematisch sind?

## Erkenntnistransfer in Lern-Allianzen

**Ziel:** Indem Organisationen transparent machen, wo KI-Systeme an Grenzen gestoßen sind oder echte Mehrwerte nur unter spezifischen Bedingungen oder hohem Aufwand erreicht werden konnten, sparen sie anderen Akteur:innen wertvolle Ressourcen und verhindern, dass Fehler wiederholt werden.

### Anleitung für ein Lernprotokoll

#### 1. Die Hypothese:

- Was wollten wir erreichen? (z.B. „Zeitersparnis bei der Moderation von Inhalten durch KI-Vorfilterung“).
- Welche Annahme hatten wir? (z.B. „Das Tool erkennt einen Verstoß gegen unsere Community-Regeln und Posts dementsprechend löschen“).

#### 2. Herausforderungen:

- Wo gab es Probleme? (z.B. „Beiträge wurden oft fälschlich als Verstoß gemeldet. Mitarbeitende mussten viel Zeit in die korrekte Prüfung investieren.“).
- Welche Auswirkungen traten auf? (z.B. „Es gab zunächst Beschwerden über gelöschte Beiträge, die den Community-Regeln entsprachen.“).

#### 3. Lösungsansätze:

- Wie haben wir das Problem gelöst? (z.B.: „Wir haben entschieden, ein System zu nutzen, das Nutzer:innen das Flaggen problematischer Posts erleichtert und Community-Manager:innen eine bessere Übersicht bietet.“)

#### 4. Empfehlungen

- Don't: (z.B. „Wir raten davon ab, zu viel Vertrauen in die automatisierte Moderation von Inhalten zu setzen.“)
- Do: (z.B. „Überlegt, ob ihr eure Nutzer:innen stärker in die Umsetzung eurer Community-Standards einbinden könnt.“)

## 03 Gemeinsame Standards für mehr Orientierung: Code of Conduct Demokratische KI

Solidarität erfordert gemeinsame Standards, geteilte Verantwortung und die Bündelung von Ressourcen, um digitale Transformationsprozesse gemeinwohlorientiert zu gestalten. In den letzten Jahren sind daher zahlreiche internationale Instrumente, Prinzipien und Deklarationen entstanden, die eine Orientierung für den verantwortungsvollen Umgang mit KI bieten.

- [Toronto Declaration \(2018\)](#):<sup>15</sup> wurde von Amnesty International und Access Now entwickelt, mit breiter Unterstützung durch zivilgesellschaftliche Akteur:innen. Sie verankert KI im menschenrechtlichen Rahmen, richtet sich aber primär an Staaten und Unternehmen.
- [Montreal Declaration \(2018\)](#):<sup>16</sup> ist in einem partizipativen Prozess der Université de Montréal mit starker Bürger:innenbeteiligung entstanden. Sie stellt ethische Prinzipien für KI in den Dienst des Gemeinwohls und betont auch explizit die Bedeutung von Solidarität.
- [Asilomar AI Principles \(2017\)](#):<sup>17</sup> Vom Future of Life Institute veröffentlicht und von über 5000 Expert:innen unterzeichnet. Sie betonen die Verantwortung für geteilten Wohlstand, moralische Werte und gesellschaftliche Kontrolle.
- [Universal Guidelines on Artificial Intelligence \(2018\)](#):<sup>18</sup> Initiiert durch die zivilgesellschaftliche Koalition The Public Voice, angegliedert an das Electronic Privacy Information Center (EPIC). Die Guidelines sind als universell anwendbare Prinzipien konzipiert.

---

<sup>15</sup>Toronto Declaration. (2018, 16. Mai). The Toronto Declaration: Protecting the right to equality in machine learning. Amnesty International & Access Now. Abgerufen am 8. Dezember 2025, von <https://www.torontodeclaration.org/declaration-text/english/>

<sup>16</sup>Montréal Declaration for a Responsible Development of Artificial Intelligence. (2018). Erklärung von Montréal für eine verantwortungsvolle Entwicklung von künstlicher Intelligenz [PDF]. Abgerufen am 8. Dezember 2025, von [https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/AL-UdeM\\_Declaration-IA-Resp\\_allemand\\_vf.pdf](https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/01/AL-UdeM_Declaration-IA-Resp_allemand_vf.pdf)

<sup>17</sup>Future of Life Institute. (2017). Asilomar AI Principles: A set of principles for beneficial AI [Online-Dokument]. Abgerufen am 8. Dezember 2025, von <https://futureoflife.org/open-letter/ai-principles>

<sup>18</sup>Center for AI and Digital Policy (CAIDP). (2018, Oktober). Universal Guidelines for AI (UGAI). Abgerufen am 8. Dezember 2025, von <https://www.caidp.org/universal-guidelines-for-ai>

Diese globalen Instrumente setzen wichtige ethische Standards. Sie bleiben jedoch oft abstrakt oder adressieren nicht direkt die spezifischen Herausforderungen von gemeinwohlorientierten Organisationen in Deutschland. Die Frage, wie über 600.000 Vereine, Stiftungen und Initiativen, von der Sozialberatung bis zur politischen Bildungsinitiative, KI verantwortungsvoll im Alltag einsetzen, erfordert daher einen spezifischen Rahmen. Doch viele Vereine und Organisationen der Zivilgesellschaft in Deutschland sind klein.<sup>19</sup> Sie verfügen selten über die Kapazität, eigene, fundierte Leitlinien für den verantwortungsvollen Umgang mit KI zu entwickeln.

Mit dem Code of Conduct Demokratische KI haben wir einen gemeinsamen Orientierungsrahmen erarbeitet, um genau diese Leerstelle zu adressieren.

### Co-Kreation als solidarischer Prozess

Im Rahmen eines vom Bundesministerium für Bildung, Familie, Senioren, Frauen und Jugend geförderten Projekts hat unter der Leitung von D64 – Zentrum für Digitalen Fortschritt ein mehr als 20-monatiger, partizipativer Prozess stattgefunden. Über 50 zivilgesellschaftliche Organisationen – darunter die Arbeiterwohlfahrt, die Amadeu Antonio Stiftung, die Deutsche Schachjugend und CorrelAid – arbeiteten in hybriden Community-Treffen und asynchroner Textarbeit zusammen. Dieser Prozess selbst war darauf ausgelegt, verschiedene Organisationen zusammenzubringen, um gemeinsame Lösungen zu erarbeiten und die Handlungsfähigkeit der Zivilgesellschaft zu stärken:

1. **Ressourcenausgleich:** Durch Reisekostenerstattungen und Aufwandsentschädigungen wurde die Teilnahme ressourcenärmerer Organisationen und ehrenamtlich Engagierter ermöglicht. Das war entscheidend, um eine breitere Vielfalt in der Gruppe zu sicherzustellen.
2. **Flexible Beteiligung:** Während die Beteiligung an vier großen Community-Treffen verpflichtend war, konnten sich Teilnehmende nach Verfügbarkeit in kleineren Arbeitssitzungen und asynchroner Zusammenarbeit in kollaborativen Dokumenten einbringen. Das ermöglichte Beteiligung auch unter den strukturellen und zeitlichen Einschränkungen zivilgesellschaftlichen Engagements.
3. **Diversität als Stärke:** Die breite Vielfalt der beteiligten Organisationen (z.B. Wohlfahrtsverbände, Umweltorganisationen, Bildungsorganisationen) sorgte für einen praxisnahen und breit anschlussfähigen Code of Conduct. Die Notwendigkeit, abstrakte Prinzipien im Kontext unterschiedlicher Anwendungsfälle zu diskutieren, ermöglichte einen gemeinsamen Lernprozess.

Das Ergebnis dieses solidarischen Gemeinschaftswerks sind acht Grundprinzipien auf zehn Seiten, die zivilgesellschaftlichen Organisationen konkrete Orientierung für den KI-Einsatz bieten. Das Bekenntnis zu gemeinsamen Prinzipien im Umgang mit KI stärkt die kollektive Haltung der Zivilgesellschaft im digitalen Diskurs.

---

<sup>19</sup>Schubert, P., Kuhn, D. & Tahmaz, B. (2023). ZiviZ-Survey 2023: Zivilgesellschaftliche Organisationen im Wandel – Gestaltungspotenziale erkennen. Resilienz und Vielfalt stärken. Hauptbericht [PDF]. ZiviZ im Stifterverband. Abgerufen am 8. Dezember 2025 von [https://www.ziviz.de/sites/ziv/files/ziviz-survey\\_2023\\_hauptbericht.pdf](https://www.ziviz.de/sites/ziv/files/ziviz-survey_2023_hauptbericht.pdf)



Insgesamt haben sich bereits über 100 Organisationen der Initiative angeschlossen. Eine zentrale Erkenntnis aus dem Prozess ist, dass die Zivilgesellschaft die digitale Transformation gemeinsam produktiv mitgestalten kann.

## Fazit

Die digitale Transformation stellt die Zivilgesellschaft vor immense Herausforderungen, doch die Gestaltungsmacht liegt in der konsequenten Verankerung gesellschaftlicher Werte. Der Code of Conduct Demokratische KI beweist, dass kollektives Handeln, die Bündelung von Ressourcen und die Etablierung gemeinsamer Standards möglich sind. Die Arbeit beginnt nun in jeder einzelnen Organisation. Solidarität zu verankern bedeutet nicht, perfekt zu sein, sondern, ehrlich zu reflektieren, Fehler zu teilen und sich aktiv für die Bedürfnisse von anderen einzusetzen.

## Mitwirkende:

Dieses Papier ist in einem internen Arbeitsprozess des Projekts **Code of Conduct Demokratische KI** unter der Leitung von **Anke Obendiek, PhD (D64)** entstanden. Mitarbeit: **Ulrich Berger (D64)**, **Laura Dederichs (Faktor D)**, **Alexander Engel (BAGSO e.V.)**, **Christoph Hassler (H:DEI)**, **Monika Ilves (D64)**, **Niels Heinemann (Kompetenzzentrum Technik - Diversity - Chancengleichheit e.V.)**, **Barbara Jung (AKTIVOLI-Landesnetzwerk Hamburg e.V.)**, **Maximilian Kühn (AWO Bundesverband e. V.)**, **Lorelei Logel-Demoulin (DRK)**, **Konstantin Macher (Digitale Gesellschaft e. V.)**, **Linda Machwitz (Digital für alle)**, **Nevena Nikolajević (Correlaid, e.V.)**, **Mario Michael Ottaiano (Weizenbaum Institut)**, **Damian Paderta (Offene Kommunen. NRW)**, **Oliver Pfeiderer (LAG Selbsthilfe RLP)**, **Dr. Jens Preußner (DRK)**, **Moritz Schmidt (LAG21)**, **Tim Schrock (AdB)**, **Adriano Schwanke (DaMOst)**, **Tarek Skouti (Common Grounds Forum)**, **Sibylle Würz (FrauenComputerZentrumBerlin e.V. (FCZB))**.

## 04 Impressum

Wir bedanken uns bei allen, die uns mit ihren Hinweisen und Kommentaren bei der Erstellung des Papiers weitergeholfen haben. Wir bedanken uns für die finanzielle Förderung des Bundesministeriums für Bildung, Familie, Senioren, Frauen und Jugend im Rahmen der Richtlinie zur Förderung von Künstlicher Intelligenz für das Gemeinwohl.

Gestaltungskonzept: [Antje Schnier](#)

Gestaltung: [Timo Knorr](#)

### Adresse

D64 – Zentrum für Digitalen Fortschritt e.V.

Co-Vorsitzender: [Erik Tuchtfeld](#)

Co-Vorsitzende: [Svea Windwehr](#)

Chausseestraße 5

10115 Berlin

### Kontakt

Anke Obendiek, PhD (D64 – Zentrum für Digitalen Fortschritt)

[buero@d-64.org](mailto:buero@d-64.org)

[demokratische-ki.de](https://demokratische-ki.de)

Dezember 2025

Gefördert vom:



Bundesministerium  
für Bildung, Familie, Senioren,  
Frauen und Jugend





## 05 Reflexionsvorlagen



## Reflexionsvorlage für die Einführung von KI-Tools

Diese Vorlage unterstützt dabei, neue digitale oder KI-gestützte Tools im Team einzuführen. Sie hilft, Auswirkungen zu reflektieren, Feedback zu strukturieren und mögliche Risiken zu erkennen.

**Zielgruppe:** Multiplikator:innen, Personen, die Tools einführen

**Dauer:** ca. 10 Minuten (1-4), danach mit Abstand von min. 1 Woche 5-6

### 1. Kurzbeschreibung des Tools

- Name des Tools
- Zweck / Einsatzziel
- Geplante Nutzung im Team / in der Organisation
- Voraussichtlicher Start der Nutzung

### 2. Erwartungen & Ziele

- Was versprechen wir uns vom Einsatz des Tools?
- Welche Herausforderungen möchten wir damit lösen?
- Welche Erwartungen haben unterschiedliche Teammitglieder?

### 3. Reflexionsfragen zur solidarischen Einführung

- Wer profitiert vom Einsatz des Tools und wer könnte benachteiligt sein?
- Gibt es Barrieren für bestimmte Gruppen (z. B. Sprache, Zugang, Technik)?
- Wurden betroffene Zielgruppen oder Kolleg:innen einbezogen?
- Welche Informationen über Herkunft, Entwicklung oder Datenlage des Tools liegen vor?
- Wie gehen wir mit Unsicherheiten, Fehlern oder Widerstand um?

### 4. Testphase & Feedbackstruktur

- Dauer der Testphase
- Wer testet?
- Welche Anwendungsfälle werden erprobt?
- Wie und wann wird Feedback gesammelt (z. B. Retrospektive, Umfrage, Gespräch)?

### 5. Gemeinsame Auswertung & Entscheidung

- Was lief gut? Was hat nicht geklappt?
- Gab es Ausschlüsse, Probleme oder unbeabsichtigte Folgen?
- Wird das Tool dauerhaft übernommen? Wenn ja, unter welchen Bedingungen?
- Welche Alternativen oder Anpassungen wurden in Betracht gezogen?

### 6. Verantwortlichkeiten & Weiterbegleitung

- Wer übernimmt die Verantwortung für das Tool (Ansprechpartner:in, Admin etc.)?
- Wie wird die Nutzung dokumentiert und regelmäßig reflektiert?
- Wann findet die nächste Überprüfung statt (z. B. in 2 Monaten)?



## Reflexionsvorlage für die Nutzung von KI-Tools

Diese Vorlage unterstützt dabei, die Nutzung neuer digitaler oder KI-gestützter Tools zu begleiten. Sie hilft, Auswirkungen zu reflektieren, Feedback zu strukturieren und mögliche Risiken zu erkennen.

**Zielgruppe:** Mitarbeitende, Ehrenamtliche oder andere Nutzer:innen

**Dauer:** ca. 20–30 Minuten

**Hinweis:** Es gibt kein „richtig“ oder „falsch“. Wichtig ist deine persönliche Erfahrung.

### 1. Basisdaten

- Name des KI-Tools:
- Wofür wurde es genutzt?  
(z. B. Textgenerierung, Chatbot, Datenanalyse ...)
- In welchem Rahmen?  
(z. B. Workshop, Kampagne, Ehrenamtskoordination)
- Wie häufig genutzt?  
(einmalig / gelegentlich / regelmäßig)

### 2. Erste Eindrücke

- Was war dein erster Eindruck vom Tool?
- Wie einfach oder schwer war der Einstieg?  
(z. B. Bedienbarkeit, Sprache, Technik)

### 3. Nutzen & Erfolge

- Was hat gut funktioniert?
- Gab es neue Erkenntnisse oder Überraschungen?
- Wurde dir etwas erleichtert? Wenn ja, was?

### 4. Herausforderungen & Grenzen

- Was hat nicht gut funktioniert?  
(z. B. falsche Ergebnisse, unklare Sprache, technische Hürden)
- Wurden bestimmte Gruppen durch das Tool ausgeschlossen oder benachteiligt?
- Gab es Situationen, in denen du lieber auf das Tool verzichtet hättest?

### 5. Soziale Wirkung

- Wie hat das Tool die Zusammenarbeit mit anderen beeinflusst?  
(z. B. Kommunikation, Vertrauen, Rollenverteilung)
- Gab es Veränderungen in der Beziehung zu Ehrenamtlichen oder Kolleg:innen?

### 6. Reflexion & Ausblick

- Was hast du durch die Nutzung gelernt?
- Würdest du das Tool weiterempfehlen? Warum (nicht)?
- Was würdest du dir für den weiteren Einsatz wünschen?  
(z. B. Schulung, Alternativen, klare Regeln)

### 7. Sonstiges

- Was möchtest du noch teilen?