



Code of Conduct
Demokratische KI

White Paper

KI und Gerechtigkeit: vier Thesen für die Zivilgesellschaft

Spannungsfelder, Praxisbeispiele und positive
Visionen



White Paper

KI und Gerechtigkeit: vier Thesen für die Zivilgesellschaft

Spannungsfelder, Praxisbeispiele und positive
Visionen

Inhalt

1. KI und Gerechtigkeit in der Zivilgesellschaft.....	2
2. Vier Thesen zu KI & Gerechtigkeit.....	5
I. Literacy und Kompetenzen	5
II. Zugang und Zugänglichkeit	9
III. Diskriminierung	12
IV. Nachhaltigkeit	16
3. Wie geht's weiter?	20
4. Impressum	22

01 KI und Gerechtigkeit in der Zivilgesellschaft

Künstliche Intelligenz (KI) erscheint oft als neutrales Werkzeug, das vermeintlich objektivere Entscheidungen verspricht. Unternehmen, Organisationen oder die öffentliche Verwaltung erhoffen sich von KI-Systemen eine faire und effiziente Verteilung begrenzter Ressourcen. Tatsächlich können KI-Systeme dazu beitragen, Chancengerechtigkeit zu verbessern oder Ungleichheiten zu dokumentieren:

- **Sprachsteuerung oder Bildererkennung** können Menschen mit Behinderungen mehr gesellschaftliche Teilhabe ermöglichen.
- **Übersetzungstools** können Sprachbarrieren abbauen.
- **Datengestützte Analysen** können Benachteiligungen auch in komplexen Zusammenhängen schneller sichtbar machen.

Was bietet das White Paper?

- Eine Analyse der Spannungsfelder zwischen Gerechtigkeit und KI.
- Vier Thesen, die den Zusammenhang von KI und Kompetenzen, Zugang, Diskriminierung und Nachhaltigkeit beleuchten.
- Praxisnahe Einblicke von zivilgesellschaftlichen Organisationen, die von Erfahrungen und Lösungsansätzen berichten.

Doch die Auswirkungen von KI sind ungleich verteilt und werfen neue Fragen von Gerechtigkeit auf. Denn in einer Welt, die von strukturellen Ungleichheiten geprägt ist, kann KI bestehende Ungerechtigkeiten nicht nur reproduzieren, sondern Machtstrukturen festigen oder verschärfen.

- Beispiele von **Diskriminierung** aus dem Personalwesen¹, der Vergabe von Sozialleistungen² oder der Strafverfolgung³ zeigen die gravierenden Konsequenzen von ungerechten KI-Systemen.
- Häufig unsichtbare menschliche Arbeit und Umweltfolgen belasten oft bereits benachteiligte Regionen.
- Die **Konzentration von Macht bei wenigen Tech-Konzernen**, die KI-Entwicklung, Infrastruktur und Daten dominieren, gefährdet **demokratische Grundwerte**. Denn diese Macht hat auch politische und gesellschaftliche Konsequenzen – ob auf öffentliche Diskurse, Bildung, den Schutz von Minderheiten oder Rechtsstaatlichkeit.

1 Dastin, J. (2018): Insight - Amazon scraps secret AI recruiting tool that showed bias against women. <https://www.reuters.com/article/us-amazon-com-jobs-automation-%20insight-idUSKCN1MK08G/>

2 Rachovitsa, A.; Johann, N. (2022): The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case, Human Rights Law Review, 22(2) ngac010. <https://doi.org/10.1093/hrlr/ngac010>

3 United Nations: General Assembly (2024): Contemporary forms of racism, racial discrimination, xenophobia and related intolerance. <https://docs.un.org/en/A/HRC/56/68>. S. 8.

Diese Komplexität kann – besonders für zivilgesellschaftliche Organisationen, die sich für mehr Gerechtigkeit einsetzen – überwältigend wirken: Wie bewerten wir gesellschaftlichen Mehrwert im Vergleich zu den negativen Konsequenzen? Kann die Zivilgesellschaft angesichts der problematischen Produktionsbedingungen, des Ressourcenverbrauchs und der Machtkonzentration bei Technologiekonzernen überhaupt einen ethisch vertretbaren Umgang mit KI finden? Und was bedeutet es für Organisationen, sich der Nutzung von KI entgegenzustellen?

Statt uns dieser Auseinandersetzung zu verweigern, plädieren wir für einen informierten und kritischen Umgang mit KI – und möchten auch positive Visionen hervorheben. Wir diskutieren vier zentrale Thesen zu KI und Gerechtigkeit: Zugang, Kompetenzen, Diskriminierung und Nachhaltigkeit. Dabei ist für uns die Frage nach Handlungsmöglichkeiten der Zivilgesellschaft zentral: Wie schaffen wir mehr Gerechtigkeit mit und im Kontext von KI?

Was ist Gerechtigkeit?

Gerechtigkeit bildet das Fundament für sozialen Zusammenhalt und steht in enger Beziehung zu einer lebendigen Demokratie. Sie umfasst die Gleichwürdigkeit und Gleichbehandlung aller Menschen genauso wie bedarfsgerechte Unterstützung und faire Ressourcenverteilung. Sie umfasst globale, generationenübergreifende und ökologische Aspekte. Im Kontext von KI wird Gerechtigkeit vor allem auf das Konzept von Fairness bezogen – also der wahrgenommenen Gerechtigkeit bei der Entwicklung oder des Einsatzes von KI-Technologien. Ohne Gerechtigkeit fehlt die Basis für die akzeptierte und legitime Umsetzung von KI-Systemen. insbesondere für zivilgesellschaftliche Organisationen, die gemeinwohlorientiert arbeiten.

Wie ist Gerechtigkeit gefährdet?

1. **Spiegelung oder Verschärfung bestehender gesellschaftlicher Ungleichheiten:** Die Planung, Entwicklung, Implementierung oder Evaluation von KI begünstigt oft privilegierte Gruppen. Darüber hinaus können Verzerrungen in Trainingsdaten oder bestimmte Wertvorstellungen der Akteur:innen zu Benachteiligungen führen.
2. **Verschärfung der „digitalen Kluft“:** Während einige gesellschaftliche Gruppen von Technologien wie KI-Systemen profitieren, wird die berufliche, soziale oder politische Teilhabe für andere erschwert – etwa durch fehlende Zugangs- oder Nutzungsmöglichkeiten oder Mangel der notwendigen Kompetenzen. Gleichzeitig werden bestehende Machtstrukturen gefestigt und gegebenenfalls verstärkt – insbesondere weil KI-Kompetenzen für die kritische Bewertung von KI-Systemen notwendig sind.
3. **Soziale Ungerechtigkeit:** Die Entwicklung von KI-Systemen basiert häufig auf unsichtbarer menschlicher Arbeit. Sogenannte Clickworker:innen kategorisieren Inhalte, bereinigen Datensätze und filtern Hassrede aus KI-Ergebnissen. Diese arbeiten häufig im globalen Süden unter ausbeuterischen Bedingungen.⁴
4. **Ökologische Gerechtigkeit:** Der hohe Ressourcen- und Energieverbrauch für KI-Infrastrukturen, etwa bei der Herstellung von Prozessoren oder beim Betreiben von Rechenzentren, belastet Ökosysteme in Regionen, die bereits unter den Folgen des Klimawandels leiden.

⁴ Kloiber, J. (2024): Putzkräfte der digitalen Plattformen: Social-Media-Content-Moderator*innen wehren sich gegen psychische Belastungen, schlechte Bezahlung und Ausbeutung. WSI-Mitteilungen 77, S. 63–65. <https://dx.doi.org/10.5771/0342-300x-2024-1-63>



Wie kann man Gerechtigkeit fördern?

Chancengleichheit: Gleiche Chancen für alle entstehen nur, wenn wir individuelle Bedarfe ernst nehmen – und sowohl auf faire Prozesse als auch auf gerechte Ergebnisse achten.

Diskriminierungskritik: Die konsequente Anwendung internationaler Menschenrechtskonventionen und die Berücksichtigung struktureller Benachteiligung sind essentiell.

Vielfalt: Diverse Perspektiven in der Entwicklung und Anwendung von KI müssen gefördert werden.

Transparenz: Durch Nachvollziehbarkeit von KI-Systemen sowie der Herkunft und Zusammensetzung von Trainingsdaten können Entscheidungen validierbar werden.

Teilhabe und **Kontrolle:** Durch echte Teilhabe werden Menschen befähigt – ob durch den Aufbau von Kompetenzen oder durch Beteiligungsstrukturen. Dabei sollten Gestaltungsrechte institutionell verankert und vor der Kontrolle durch bestimmte Gruppen geschützt werden.

Regelmäßige Überprüfung: Gerechtigkeitsvorstellungen unterliegen ständigen gesellschaftlichen Aushandlungsprozessen. Daher verlangt auch gerechter KI-Einsatz kontinuierliche Reflexion, um Missstände zu erkennen und auf diese zu reagieren.

02 Vier Thesen zu KI & Gerechtigkeit

I. Literacy und Kompetenzen

KI-Systeme sind mit vielfältigen Erwartungen verknüpft – von der Hoffnung auf Weltrettung bis hin zur Angst vor existenziellen Bedrohungen. Schon heute beobachten wir bei digitalen Geräten und Anwendungen eine digitale Kluft: Auf der einen Seite stehen informierte und kompetenten Nutzer:innen, auf der anderen jene, die technische Systeme zwar in unterschiedlicher Form nutzen, aber weder deren

Funktionsweise noch deren mögliche Auswirkungen oder die Interessen der Anbieter verstehen. Diese Entwicklung wird durch KI-Systeme – und insbesondere die automatische Integration von KI in bestehende Tools – noch verstärkt.⁵ Statt sich diffusen Ängsten oder übersteigerten Erwartungen hinzugeben, hilft „KI-Kompetenz“ –im Folgenden als „AI Literacy“ bezeichnet – bei der kritischen Bewertung und Einordnung von KI-Systemen.

„KI-Kompetenzen sind die Basis um KI und Gerechtigkeit in Einklang zu bringen“

AI Literacy

AI Literacy umfasst Kompetenzen, die Nutzer:innen einen analytisch-kritischen Umgang mit KI-Technologien sowohl auf der Reflexions- als auch auf der Anwendungsebene erlauben:

- Verständnis der **technischen Grundlagen**,
- kritische Auseinandersetzung mit **Chancen und Grenzen**,
- Verständnis **ethischer und rechtlicher Aspekte**,
- **effektive Nutzung** von KI-Anwendungen.

Dabei ganz zentral: KI ist kein Allheilmittel oder Selbstzweck, sondern ein Werkzeug – oder vielmehr eine Palette verschiedener Werkzeuge für verschiedene Aufgaben. Die entscheidende Frage lautet nicht, ob KI generell nützlich ist, sondern ob das spezifische Werkzeug zur vorliegenden Aufgabe passt: Ob wir einen Nagel einschlagen, eine Schraube festziehen oder einen Ast sägen wollen, entscheidet darüber, welches Werkzeug wir benutzen – und wie das Ergebnis aussieht.

5 Wienrich, C.; Carolus, A.; Markus, A., Augustin, Y. (2022): AI Literacy: Kompetenzdimensionen und Einflussfaktoren im Kontext von Arbeit. BMAS Working Paper. https://www.denkfabrik-bmas.de/fileadmin/Downloads/Publikationen/AI_Literacy_Kompetenzdimensionen_und_Einflussfaktoren_im_Kontext_von_Arbeit.pdf



Gleiches gilt für den Einsatz von KI. Die Auswahl und Bewertung des passenden Werkzeugs hängt dabei von Entwicklung, Betrieb und Funktionsweise der KI-Systeme ab – nicht von einer pauschalen Entscheidung für oder gegen KI.

Vorüberlegungen bestimmen den Prozess

Vor dem Einsatz von KI bedarf es einer grundsätzlichen Abwägung der möglichen Vor- und Nachteile von KI. Oftmals ist es effizienter, bestehende Prozesse kritisch zu hinterfragen und neu zu gestalten, statt einfach KI zu integrieren. Denn strukturelle Probleme oder schlecht laufende Prozesse erfordern grundsätzliche Änderungen. Für einen sinnvollen Einsatz von KI ist es entscheidend, bereits vor der Einführung konkrete Ziele und eine klare Zielgruppe zu definieren.

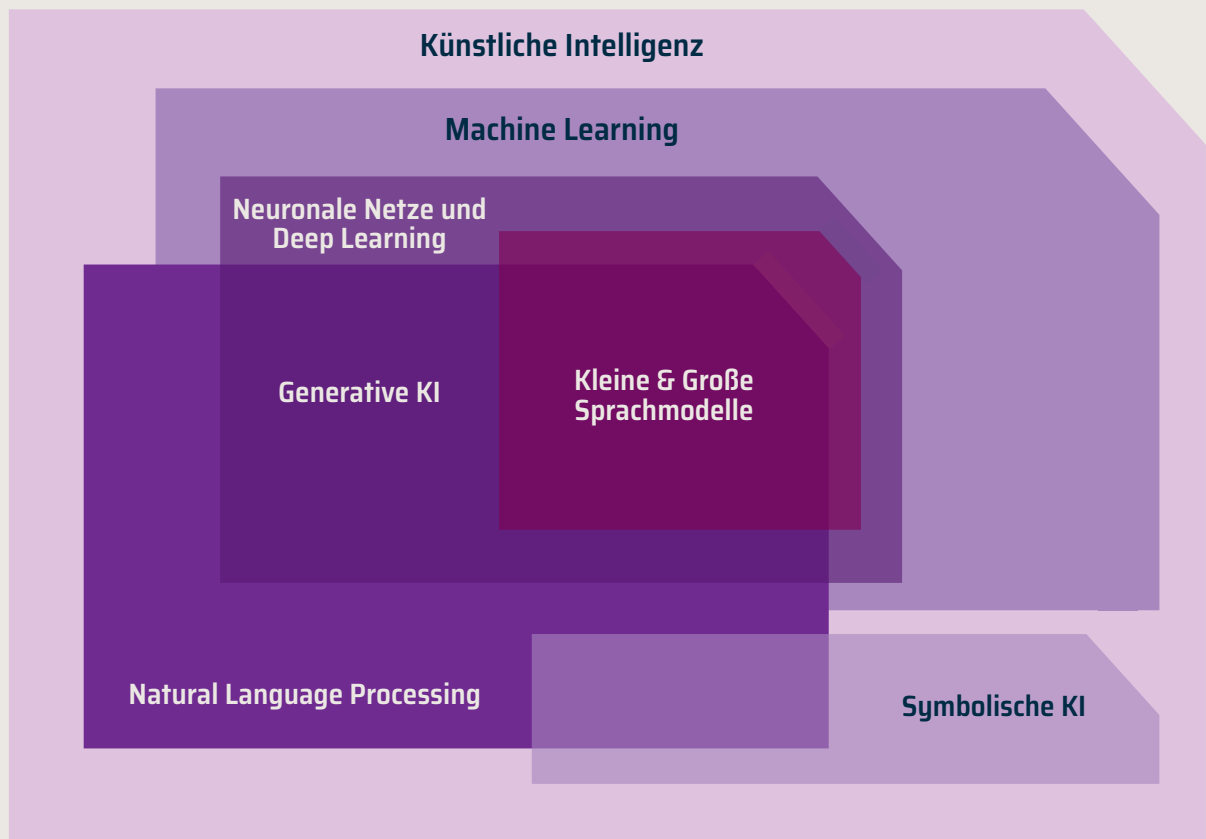
Wie groß ist der Kreis der Nutzenden und der Beteiligten? Welcher Aufwand und welche Mehrwerte sind mit einer Implementierung verbunden? Kann ein langfristiger Betrieb sichergestellt werden? Gibt es eine ressourcensparende Alternativlösung?

KI ist vielfältig – und mehr als generative KI

Die aktuelle Aufmerksamkeit um KI basiert vor allem auf Fortschritten im Bereich von sogenannten neuronalen Netzwerken, deren theoretische Grundlagen bis in die 1940er Jahre zurückreichen. Erst die exponentiell gestiegene Rechenleistung moderner Hardware ermöglicht heute deren massentaugliche Anwendung.

Doch KI umfasst mehr als neuronale Netzwerke. Im Bereich des maschinellen Lernens, also der Fähigkeit aus Daten zu lernen, existieren etwa Klassifikationsbäume, deren Entscheidungspfade im Vergleich zu neuronalen Netzwerken leichter nachvollziehbar sind. Im Gegensatz dazu basiert die sogenannte symbolische KI auf vordefinierten Regeln oder Logikmustern statt auf Trainingsdaten. Das darauf basierende KI-Modell „Deep Blue“ schlug bereits Ende der 90er Jahre den damaligen Schach-Weltmeister.

KI lässt sich auch nach abgebildeten Fähigkeiten unterscheiden. Der Bereich des „Natural Language Processing“ (NLP) konzentriert sich auf Sprachwiedergabe. Modelle, die selbst Daten, also neben Text auch etwa Bilder oder Videos, produzieren können, werden „generative“ KI genannt. Auch die Modellgröße variiert erheblich: von einigen Millionen bis zu mehreren Milliarden Parametern.



Um hier eine Differenzierung zu ermöglichen, zeigt die Grafik einen Ausschnitt der KI-Landschaft: Gesellschaft für Informatik, 2025, abgerufen am 04. März 2025: <https://nextcloud.gi.de/s/S84BFXBMkg8ks9E>. Zur weiteren Vertiefung siehe auch Digital für alle (2025): Künstliche Intelligenz entdecken und verstehen, abgerufen am 28. April. <https://digitaltag.eu/ki-kurse#kurs1>

Data Literacy als Grundlage für AI Literacy

Nevena Nikolajević, CorrelAid e.V.

Grundlage für einen reflektierten Umgang mit KI-Technologien sind grundlegende Datenkompetenzen. Data Literacy (und Digital Literacy) bilden die Basis für AI Literacy. Da häufig bereits grundlegende Datenkompetenzen fehlen, sollte die Kompetenzvermittlung bereits hier ansetzen.

Als CorrelAid e.V. widmen wir uns genau dieser Vermittlung von grundlegenden Datenkompetenzen. Im Rahmen unserer Data Literacy-Kurse lernen zivilgesellschaftliche Akteur:innen unter anderem, Datenquellen in der eigenen Organisation zu identifizieren und die Daten hinsichtlich ihrer Nutzungspotenziale zu bewerten. Diese Fähigkeit ist für einen differenzierten Umgang mit KI-Anwendungen unerlässlich.

Will man eine eigene KI-Anwendung entwickeln oder eine bestehende an die eigenen Bedarfe anpassen, benötigt man Daten – viele Daten, die einem gewissen qualitativen Niveau entsprechen. Diese sollten etwa in maschinenlesbarer Form und in einem einheitlichen Schema vorliegen. Eine gute Datenkultur sollte daher im besten Fall bereits in der Organisation etabliert sein oder zumindest parallel zur Nutzung von KI-Anwendungen etabliert werden. Ein Bericht der Organisation Data Orchard zeigt jedoch, dass der Aufbau von Datenkompetenz für über 75% der befragten Organisationen im Nonprofit-Sektor eine Herausforderung darstellt.⁶

Die „richtigen“ Fragen stellen

Wird die Entwicklung oder Beauftragung einer KI-Anwendung geplant, sollte abgeschätzt werden können, welche Aspekte im konkreten Fall zu berücksichtigen sind. Auf welchen Daten wurde ein Modell trainiert bzw. soll es trainiert werden? Besteht eine Verzerrung (d.h. wurden Vorurteile in das Modell eingespeist) und wie kann diese vermieden werden? Werden personenbezogene Daten verwendet? Welche langfristigen Effekte sind durch den Einsatz der Anwendung zu erwarten?

Es ist nicht notwendig, die Bewertung der verschiedenen Aspekte vollständig selbst vorzunehmen. Man sollte aber die richtigen Fragen stellen und einschätzen können, zu welchen Aspekten zusätzliche Expertise notwendig ist.

Im Rahmen des Civic Data Lab, einem gemeinsamen Vorhaben der Gesellschaft für Informatik e.V., CorrelAid e.V. und dem Deutschen Caritasverband e.V., wird derzeit beispielsweise ein E-Learning zum Thema „Datenmanagement und -infrastruktur“⁷ entwickelt, das viele relevante Konzepte behandelt, die grundlegend sind, um die Funktionsweise und den Betrieb von KI-Anwendungen zu verstehen.

⁶ Data Orchard (2024): State of the Sector: Data Maturity in the Nonprofit Sector 2024: <https://www.dataorchard.org.uk/resources/sots-data-maturity-in-nonprofit-sector-2024>

⁷ Civic Data Lab (2025): Selbstlernmaterial. <https://civic-data.de/selbstlernmaterial/>

II. Zugang und Zugänglichkeit

„Der ungleiche Zugang zu KI schafft neue digitale Klüfte, die wir überwinden müssen.“

Die Gestaltung des Zugangs zu KI entwickelt sich zu einer zentralen Gerechtigkeitsfrage unserer Zeit. Während KI-Systeme immer mehr Lebensbereiche durchdringen, zeigt sich eine wachsende Kluft zwischen denen, die Zugang zu diesen Technologien haben und sie nutzen können, und denen, die davon ausgeschlossen bleiben.

Der Zugang zu KI umfasst verschiedene Dimensionen:

- Der **technische Zugang** bildet die Grundvoraussetzung – von der notwendigen Hardware bis hin zu KI-Anwendungen.
- **Finanzielle Hürden** können den Zugang zu leistungsfähigen KI-Anwendungen erschweren.
- Die **Qualität der Infrastruktur** ist ebenfalls entscheidend – ob schnelle und stabile Internetverbindungen, der Zugang zu Datensätzen oder Cloud- und Serverkapazitäten.
- **Sprachliche und kulturelle** Aspekte beeinflussen die Zugänglichkeit, etwa durch fehlende linguistische Diversität in Training und Oberflächen von Sprachmodellen oder mangelnder Berücksichtigung des sozialen Kontexts.

Für gemeinwohlorientierte Organisationen stellt sich die Zugangsfrage in doppelter Weise: Einerseits müssen sie ihren eigenen Zugang zu KI-Technologien sicherstellen, andererseits die Interessen ihrer Zielgruppen vertreten. Die ungleiche Verteilung von Ressourcen trifft dabei besonders kleine Initiativen, die mit begrenzten Budgets, fehlender technischer Expertise und knappen personellen Ressourcen kämpfen. Die zunehmende Konzentration technologischer Kapazitäten bei wenigen großen Akteuren verstärkt dabei bestehende Machtungleichgewichte. Die Zivilgesellschaft steht daher vor der Herausforderung, eine Balance zwischen ihren Werten und praktischer Umsetzbarkeit zu finden.

Nur wenn wir die Gestaltung gerechter Zugangsmöglichkeiten als gesamtgesellschaftliche Herausforderung begreifen, können wir verhindern, dass sich bestehende soziale Ungleichheiten durch diese Technologie weiter verschärfen. Wie der Zugang zu KI-Systemen in Zukunft gestaltet sein wird, hängt maßgeblich davon ab, wie gut es gelingt, technologische Innovation mit gesellschaftlicher Verantwortung zu verbinden.

Wie können wir Zugangsbarrieren überwinden?

Zugangsbarrieren wirken oft nicht isoliert, sondern verstärken sich gegenseitig. Nur durch das Zusammenspiel verschiedener Maßnahmen und Kooperation können gemeinwohlorientierte Organisationen Zugangsbarrieren nachhaltig überwinden.

- **Technische und strukturelle Barrieren:** Herausforderungen wie mangelnde technische Infrastruktur, Stadt-Land-Gefälle, ungleiche Ressourcenverteilung und fehlende Unterstützungsstrukturen erfordern Kooperationen. Neben der Nutzung ressourcenschonender Anwendungen oder der Entwicklung von offline-fähigen Systemen können strategische Allianzen durch die gemeinsame Nutzung von Lizenzen oder technischer Infrastruktur Kosten senken und Daten zugänglich machen.
- **Sprachliche und kulturelle Barrieren:** Die Dominanz englischsprachiger KI-Systeme und fehlende lokale Anpassungen können die Zugänglichkeit von KI verringern. Sprachmodelle können etwa bei bestimmten Dialekten negative Stereotype reproduzieren⁸, Diversität in der Entwicklung, Anwendung und Evaluation von KI, die Ausrichtung auf die Bedarfe von Nutzenden sowie mehrsprachige Benutzeroberflächen ermöglichen Zugang für unterschiedliche Gruppen.
- **Kompetenzbezogene Barrieren:** Fehlende Expertise, mangelndes Verständnis und Unsicherheit bei der Bewertung von KI-Systemen sowie Unklarheiten über rechtliche Rahmenbedingungen können Zugangsbarrieren darstellen (siehe oben). Durch Wissenstransfer zwischen Organisationen und Schulungsangebote, Rechtsberatung sowie klare Verantwortlichkeiten und interne Leitlinien können gemeinwohlorientierte Organisationen komplexen Anforderungen begegnen.

KI-Systeme an der Zivilgesellschaft ausrichten

Durch den gezielten Aufbau von Kompetenzen, die Schaffung gemeinsamer Netzwerke und die Entwicklung gemeinsamer Positionen können zivilgesellschaftliche Organisationen eine aktivere Rolle bei der Gestaltung von Zugangsmöglichkeiten von KI-Anwendungen einnehmen. So entsteht gemeinsam ein Gegengewicht zur zunehmenden Tech-Monopolisierung. Durch gemeinsame Lösungen können zivilgesellschaftliche Organisationen Ressourcen bündeln und passgenaue Anwendungen entwickeln, die datenschutzkonform und partizipativ funktionieren. Beispielsweise müssen Daten nicht auf den Servern großer Konzerne liegen und durch alternative Infrastrukturen oder Plattformen können sogenannte Lock-in-Effekte vermieden werden, die zu Abhängigkeiten von einzelnen Anbietern führen. Darüber hinaus ermöglichen offene Standards und geteiltes Wissen die Investition in öffentliche Güter. Durch den kontinuierlichen Austausch innerhalb der Zivilgesellschaft können alle von den Erfahrungen derjenigen Organisationen profitieren, die wegen ihrer Größe, Mittel oder Expertise bereits KI-Anwendungen entwickeln oder implementieren.

„KI-Systeme können auch ohne breiten Zugang zu mehr Gerechtigkeit beitragen, wenn sie gezielt eingesetzt werden.“

⁸ Hofmann, V., Kalluri, P.R., Jurafsky, D. et al. (2024): AI generates covertly racist decisions about people based on their dialect. Nature 633, 147–154. <https://doi.org/10.1038/s41586-024-07856-5>

Digitale Spracheingabe in der Pflegedokumentation

Ruven Börger, ASB Deutschland e.V.

Pflegekräfte dokumentieren täglich bis zu zwei Stunden lang ihre Arbeit, häufig zunächst per Notiz auf dem Handrücken. Hier kann Spracheingabe echte Entlastung schaffen. Der ASB setzt digitale Lösungen gezielt dort ein, wo sie mehr menschliche Nähe ermöglichen. In einem Modellprojekt hat der Verband überprüft, ob Pflegekräfte diese positiven Effekte bestätigen können.

Spracheingabe ermöglicht es Pflegekräften, Dokumentationen direkt vor Ort oder unterwegs zu erfassen. Eine App transkribiert die gesprochenen Worte in Text und integriert sie automatisch an der richtigen Stelle im bestehenden Dokumentationssystem. So können beispielsweise Vitalwerte oder Wunddokumentationen effizient und präzise festgehalten werden.

Der ASB-Bundesverband stellte stationären und ambulanten Pflegeeinrichtungen dafür eine Förderung über 12 Monate zur Verfügung und schloss eine Kooperation mit dem Technologiepartner voize. Von dreißig interessierten Einrichtungen startete zum Projektstart zunächst nur eine einzige. Mittlerweile haben zehn Einrichtungen die Software eingeführt und ein Landesverband ist dabei, sie in allen ambulanten Diensten einzuführen.

Denn die Ergebnisse des Pilotprojekts sind vielversprechend: Die Dokumentationsqualität hat sich deutlich verbessert, und die Pflegekräfte fühlen sich in ihrer Arbeit entlastet.⁹ Zudem baut die Software Schreibhürden ab und funktioniert auch bei Dialekten oder Akzenten, was die Akzeptanz und Nutzer:innenfreundlichkeit der App erhöht. Schulungen und ein Testmodus geben den Pflegekräften Sicherheit im Umgang mit der Technologie. Eine Zeitersparnis konnte und sollte im Rahmen des Projektes nicht erhoben werden. Eine Betriebsvereinbarung schließt eine Arbeitsverdichtung durch die Spracheingabe aus.

Ein wichtiger Erfolgsfaktor für die Einführung der App sind Kapazitäten vor Ort und die Förderung sowie Vorbereitung inklusive der Bewertung verfügbarer Technologien, die eine Beteiligung ermöglichen. Dr. Uwe-Martin Fichtmüller, ASB-Hauptgeschäftsführer des Bundesverbandes, betont: „Digitalisierung soll Mitarbeitenden im ASB mehr Zeit geben, sich auf das Wesentliche in ihrem Job zu konzentrieren.“

⁹ ASB (2024): Evaluation mit Pflegekräften in einem Pflegedienst: N = 43, Beteiligung an der Abfrage liegt bei 60 Prozent

III. Diskriminierung

„KI reproduziert häufig strukturelle Diskriminierung – dem müssen wir aktiv entgegenwirken.“

Die Nutzung von KI wirft wichtige Fragen zur Diskriminierung auf. In den vergangenen Jahren haben sich Schlagzeilen über die problematischen Entscheidungen von KI-Systemen gehäuft: Die Google Photos-App klassifizierte Schwarze Menschen als Gorillas, eine HP-Software zur Gesichtserkennung versagte bei Schwarzen Nutzer:innen, und Google Translate

verstärkte Geschlechterstereotype, indem es etwa „Nurse“ automatisch als weiblich und „Doctor“ als männlich übersetzte. Systeme, die in Verwaltungen und Arbeitsämtern eingesetzt wurden, diskriminierten Bedürftige, wenn es um die Zusage von Hilfsmaßnahmen ging.¹⁰

Um zu verstehen, wie es zu solcher Diskriminierung kommt, hilft es, zwei zentrale Kategorien von maschinellem Lernen zu verstehen:

1. Das überwachte Lernen basiert auf einem gesteuerten Prozess mit klar gelabelten Daten – etwa durch CAPTCHA-Tests, bei denen Menschen Bilder von Brücken oder Ampeln als solche markieren.
2. Beim unüberwachten Lernen suchen Algorithmen selbstständig nach Mustern oder Beziehungen in Daten, wobei die Entstehung der Ergebnisse oft schwer nachvollziehbar ist – und so auch Diskriminierung.

Beide Arten benötigen Trainingsdaten, die unsere Gesellschaft mit all ihren Vorurteilen und Ungerechtigkeiten widerspiegeln. Wenn etwa Personaldaten zur Auswahl von Bewerber:innen genutzt werden, werden bestehende Ungleichgewichte in der Belegschaft fortgeschrieben.

Diskriminierung und KI

KI-Systeme können gesellschaftliche Ungleichheiten nicht nur reproduzieren, sondern auch verstärken. Dies betrifft vor allem zwei Ebenen von Ungleichheiten:

1. **Verzerrungen im Output:** KI-Systeme können diskriminierende Ergebnisse produzieren, etwa aufgrund von verzerrten Trainingsdaten.
2. **Fehlender Zugang:** Die Planung, Entwicklung, Implementierung und Evaluation von KI-Systemen liegt häufig in den Händen bestimmter (privilegierter) Gruppen (siehe These II zu Zugang oben). Dies kann dazu führen, dass Verzerrungen weniger sichtbar sind und Machtungleichheiten verstärkt werden.

Viele große KI-Modelle basieren auf Quellen wie englischsprachigen Zeitungsarchiven und Online-Foren wie Reddit – Plattformen, die häufig von weißen Männern geprägt sind. Diese einseitige Datenbasis führt zu verzerrten Ergebnissen. Verschärft wird dieses Problem durch die mangelnde

¹⁰ für weitere Beispiele, siehe Orwat, C. (2019): Diskriminierungsrisiken durch Verwendung von Algorithmen. https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/studie_diskriminierungsrisiken_durch_verwendung_von_algorithmen.pdf?__blob=publicationFile&v=3

Vielfalt in den Entwicklungsteams. Dadurch werden problematische Ergebnisse oft nicht als solche erkannt – wie bei Gesichtserkennungssystemen, die Schwierigkeiten haben, Menschen mit dunkler Haut zu identifizieren.

Bias

Der Begriff Bias beschreibt hier systematische Verzerrungen, die zu Fehleinschätzungen und ungerechten Entscheidungen führen können. Im Bereich von KI gibt es dafür verschiedene Ursachen:

- **Trainingsdaten** spiegeln die Praxis und Datenlage von Organisationen oder (profitorientierter) Unternehmen und damit **bestehende gesellschaftliche Ungleichheiten** wider. Fehlende oder fehlerhafte Daten führen daher zur Verzerrung der Datenbasis und damit der Lerngrundlage für beispielsweise selbstlernende Algorithmen.
- **Algorithmen oder Systeme** können so angelegt sein, dass sie Ungleichbehandlung verstärken, etwa durch Unterschiede in der Gewichtung bestimmter Merkmale oder durch Interaktion. Auch wenn Systeme in neuen Kontexten eingesetzt werden, kann dies zu Verzerrungen führen.
- **Proxy-Diskriminierung**: Scheinbar neutrale Merkmale wie die Postleitzahl des Wohnorts können so genannte Proxy-Variablen darstellen, die vermeintlich Rückschlüsse über Personen erlauben – zum Beispiel ihr Geschlecht, ihr Alter oder ihre Herkunft. So können auch ohne Zugriff auf sensible Daten diskriminierende Ergebnisse entstehen
- Durch den sogenannten **Automation Bias** neigen Menschen außerdem dazu, den Aussagen von Maschinen eher zu vertrauen. Das kann dazu führen, dass sie diskriminierende Ergebnisse weniger kritisch hinterfragen.

Durch Biases werden häufig vulnerable Gruppen oder Menschen, die aufgrund mehrerer Merkmale Diskriminierung erfahren, benachteiligt. Etwa durch . Diese Benachteiligungen, etwa beim Zugang zu Sozialleistungen oder stereotype Darstellungen in KI-generierten Bildern, sind häufig nur schwer nachzuvollziehen oder nachzuweisen. Denn KI-Systeme sind meist komplex oder intransparent. Die KI-Verordnung der europäischen Union (z.B. Artikel 5)¹¹ erkennt diese Diskriminierungsrisiken an und verbietet beispielsweise die Ausnutzung von Schutzbedürftigkeit von Personen und Gruppen oder die Klassifizierung und Bewertung sozialen Verhaltens. Dennoch sind Diskriminierungsbetroffene bisher nicht ausreichend durch rechtliche Regelungen, wie das Allgemeine Gleichbehandlungsgesetz oder die Datenschutzgrundverordnung geschützt.¹²

11 Verordnung (EU) 2024/1689 über künstliche Intelligenz vom 13. Juni 2024, ELI: <http://data.europa.eu/eli/reg/2024/1689/oj>

12 Spiecker gen. Döhmman, Indra und Towfigh, Emanuel V. im Auftrag der Antidiskriminierungsstelle des Bundes (2023): https://www.antidiskriminierungsstelle.de/SharedDocs/forschungsprojekte/DE/RG_AGG_u_Schutz_v_Diskr_d_Algorithmen.html?nn=305536#comment



Wie können wir in der Zivilgesellschaft Diskriminierung durch KI minimieren?

Eigene Maßnahmen gegen Diskriminierung und Bias für gemeinwohlorientierte Organisationen sind zentral:

- 1. Klare Leitlinien und Erklärbarkeit:** KI-Systeme sollten so nachvollziehbar und fair gestaltet sein wie möglich –, in der Zivilgesellschaft brauchen wir Leitlinien, die Fairness, Transparenz und Erklärbarkeit in den Fokus rücken. Nur mit der notwendigen Sensibilisierung und dem Aufbau von Kompetenzen können diese auch gemeinsam umgesetzt werden.
- 2. Kontinuierliche Aufsicht, Evaluation und Kontrolle:** Klare Verantwortlichkeiten und die Überwachung von Standards sind notwendig für den sinnvollen Einsatz von KI-Systemen. Je einschneidender die Folgen der KI-Anwendungen, desto umfassender sind auch die Schutzbedürfnisse von Beteiligten und die Notwendigkeit zur Anfechtbarkeit von Entscheidungen.
- 3. Partizipative Gestaltung:** Menschliche Aufsicht allein reicht nicht aus, um Diskriminierungen zu minimieren. Potenziell betroffene Gruppen sollten aktiv in die Entwicklung, Implementierung und Evaluierung von KI einbezogen werden. Je nach Sektor und Kontext können demokratische Aushandlungsprozesse erforderlich sein, um vielfältige Perspektiven zu berücksichtigen.

„Der gezielte Einsatz von KI kann Menschen mit Diskriminierungserfahrungen stärken und unterstützen.“

Mit KI-Systemen Gleichberechtigung und Teilhabe fördern

Christoph Hassler, H:DEI

Obwohl KI häufig durch diskriminierende Tendenzen auffällt, birgt sie auch Potential zur Förderung von Gleichberechtigung und Teilhabe. Besonders in sozialen Einrichtungen kann sie zur Vereinfachung von Verwaltungsabläufen beitragen. Ein innovatives Beispiel hierfür ist das von der Caritas entwickelte CariFIX, das durch intelligente Vernetzung den fachlichen Austausch zwischen Beratenden optimiert. Solche technologischen Ansätze könnten künftig auch benachteiligten Gruppen dabei helfen, sich besser zu organisieren. Insbesondere bei der Schaffung von Barrierefreiheit kann sich KI als wertvoll erweisen: Sehbehinderte Menschen können durch Vorlesefunktionen und digitale Assistenzsysteme unterstützt werden, während Spracherkennungstechnologie Menschen mit motorischen Einschränkungen neue Zugangsmöglichkeiten zu digitalen Geräten eröffnet. Die Technologie schafft zudem wichtige Hilfestellungen für Menschen mit Hörbeeinträchtigungen, Leseschwächen oder geringen Sprachkenntnissen – etwa durch KI-gestützte Untertitelung und Übersetzung.

Das Startup SummAI bietet ein Tool zur automatischen Erstellung von Texten in „Leichter Sprache“ zur Verwendung für Organisationen oder die eigene Website an. Das Unternehmen nutzt dabei ein auf einem LLM basierendes Tool zur Vereinfachung der Sprache, sodass Inhalte auch für Menschen verständlich formuliert werden, deren sprachliche oder kognitive Fähigkeiten unterschiedliche Einschränkungen haben. Für Leichte Sprache gelten feste Regelwerke: Texte müssen mit erklärenden Bildern oder Grafiken versehen und von mindestens zwei betroffenen Menschen geprüft werden. Tools wie SummAI können daher bei der Vereinfachung von Sprache unterstützen, erfordern jedoch weitere Arbeitsschritte und die Einbeziehung von Prüfgruppen. Auch in anderen Bereichen können KI-Anwendungen menschliche Arbeit ergänzen, aber nicht ersetzen. Der LLM-basierte Chatbot YOUNA etwa unterstützt Menschen, die Rassismus erfahren. Er bietet einen geschützten Raum, in dem Diskriminierungsbetroffene ihre Erlebnisse mit Alltagsrassismus reflektieren und einordnen können – ohne sich dem Urteil anderer aussetzen zu müssen. Angesichts der unmittelbaren Auswirkungen solcher KI-Systeme auf vulnerable Gruppen ist bei ihrer Entwicklung und Implementierung besondere Sorgfalt erforderlich. Es bleibt daher eine zentrale Aufgabe, KI-Systeme regelmäßig kritisch zu hinterfragen und ihre Wirkung zu überprüfen.

IV. Nachhaltigkeit

Die verantwortungsvolle Nutzung von KI erfordert auch eine Auseinandersetzung mit dem Thema Nachhaltigkeit: Denn KI-Systeme können zwar dabei helfen, Ressourcen effizienter zu nutzen, doch sowohl ihr Training als auch ihr Betrieb verursachen erhebliche Umweltbelastungen.

- **Unverhältnismäßiger Anstieg des Ressourcenbedarfs:** Der Energieverbrauch großer KI-Modelle ist beträchtlich. Die Internationale Energieagentur schätzt, dass Rechenzentren, Kryptowährungen und KI bereits 2022 ca. 2% des globalen Strombedarfs ausgemacht haben. Bis 2026 wird ein zusätzlicher Energiebedarf von 160 TWh bis 590 TWh erwartet – dies entspricht etwa dem Energiebedarf Schwedens oder Deutschlands.¹³
- **Rohstoffe und Elektronikschrott:** Leistungsfähige Chips und andere Hardware verursachen durch den Abbau der verwendeten Rohstoffe, etwa Kobalt oder Palladium, oder den Wasserverbrauch bei der Herstellung zusätzliche Umweltschäden. Auch die wachsenden Mengen an Elektronikschrott können Probleme für lokale Gemeinschaften erzeugen.
- **KI-Industrie verschärft die Wasser-Krise:** Da Wasser für die Kühlung von Servern benötigt wird, führt der Ausbau von Serverfarmen vor allem in Regionen, in denen bereits Wasserknappheit herrscht, zu einer Verschärfung der Situation.
- **Globale Ungleichverteilung von Umweltkosten:** Während die Vorteile von KI-Systemen hauptsächlich den Ländern zugutekommen, die ohnehin eine bereits gut ausgebaute technologische Infrastruktur haben, tragen ärmere Regionen überproportional die ökologischen Lasten.

Auch, wenn es gelingt, den Ressourcenverbrauch zu reduzieren, bleiben grundlegende sozial-ökologische Herausforderungen, zu denen neben der starken Marktkonzentration von KI-Unternehmen auch prekäre Arbeitsbedingungen beim Clickworking gehören.¹⁴ Daher ist eine gesellschaftliche Debatte darüber, wann der Einsatz von KI trotz der Kosten gerechtfertigt ist, von zentraler Bedeutung.

„Durch gemeinsame Nutzung und Bereitstellung von KI-Infrastrukturen können wir Ressourcen sparen und Gemeinwohlorientierung verankern.“

¹³ International Energy Agency (2024): Electricity. <https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf>, S. 31

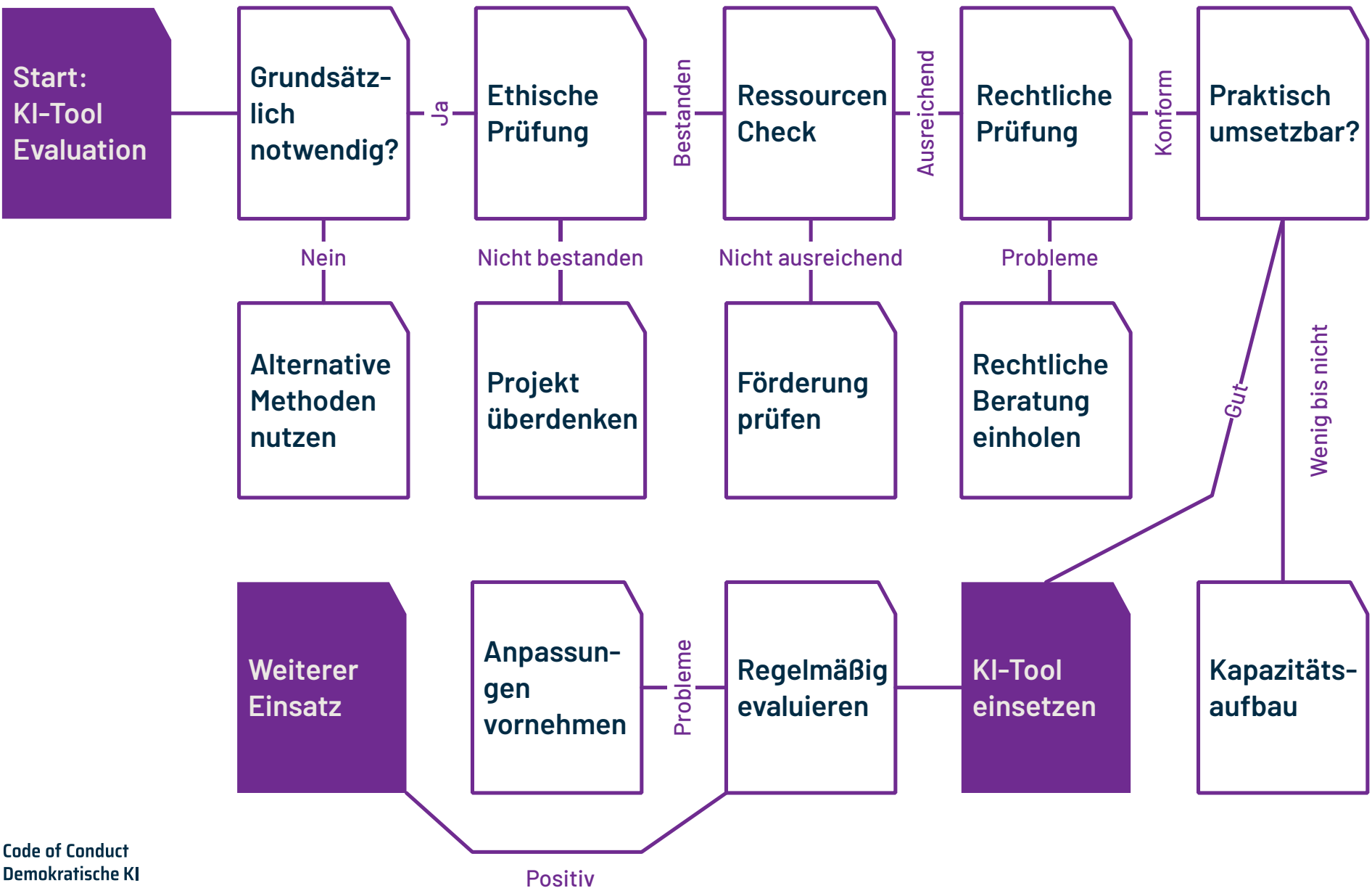
¹⁴ Civic Data Lab (2024): Clickwork – unsichtbare Arbeit hinter der 'künstlichen Intelligenz'. <https://civic-data.de/click-work-unsichtbare-arbeit/>



Wie können wir Nachhaltigkeit unterstützen?

- **Transparenz einfordern:** Es braucht mehr öffentliche Informationen über den Ressourcenverbrauch von KI-Systemen. Die KI-Verordnung führt, etwa für große Sprachmodelle, Rechenschaftspflichten und Maßnahmen zur Risikominimierung ein. Doch häufig berücksichtigen diese nicht den Energie- oder Wasserverbrauch bei der Verwendung, etwa durch Serverkapazitäten oder zu Kühlung. Nutzer:innen fehlen daher verlässliche Daten zur Nachhaltigkeit und Klimaschädlichkeit von Modellen.
- **Infrastruktur mitdenken:** Der Diskurs über KI-Anwendungen sollte auch eine Betrachtung der notwendigen „Hardware“ berücksichtigen. Denn KI verbraucht nicht nur beim Training oder bei der Verwendung Energie, sondern benötigt auch umfassende digitale Infrastruktur.
- **Kleinere Modelle nutzen:** Zum Teil können Systeme energieeffizienter genutzt werden, etwa durch Quantisierung – eine Art Kompression, um die Komplexität von Signalen zu verringern. Dies kann den Speicherbedarf und die Größe eines Modells deutlich verringern, um Modelle auch ohne große Serverkapazitäten zu nutzen. Durch die Verwendung kleinerer, spezialisierter Modelle können viele Aufgaben weniger ressourcenintensiv und passgenauer gelöst werden.
- **Nutzung abwägen:** Je nach Problemstellung können andere Methoden durchaus effizienter und genauer sein. Bei der Entscheidung über den Einsatz einer Methode, insbesondere für gemeinwohlorientierte Organisationen, sollten die sozial-ökologischen Dynamiken rund um die Technologie reflektiert werden. Denn oft stehen große profitorientierte Unternehmen hinter den übersteigerten Erwartungen in KI-Systeme. Die endgültige Entscheidung sollte immer davon abhängen, ob die positiven oder negativen Auswirkungen überwiegen.

Flowchart/Entscheidungs-



ParrotPark: Ein Prototyp im Projekt Code of Conduct Demokratische KI

Jonas Stettner, CorrelAid e.V.

Die Zusammenarbeit von zivilgesellschaftlichen Organisationen kann der Schlüssel zu einer gerechteren und nachhaltigeren Nutzung von KI-Systemen sein. Im Projekt Code of Conduct Demokratische KI haben wir daher die Entwicklung von ParrotPark angestoßen. ParrotPark ist ein Prototyp eines datenschutzfreundlichen, kontrollierbaren KI-Chatbots. Ziel ist es, zu testen, ob sich typische Anwendungsfälle von LLMs in der Zivilgesellschaft mit dem Einsatz von offenen Modellen, begrenzten GPU-Ressourcen und dem Anspruch, Modelle komplett selbst zu betreiben, umsetzen lassen. Den GPU-Bedarf – die notwendige Rechenleistung von Grafikprozessoren – kann reduziert werden, etwa durch die Verwendung von kleineren Modellen und zusätzlicher Kompression durch Quantisierung – der Verringerung der Anzahl an Bits zur schnelleren Verarbeitung. Das Vorhaben nutzt Open-Source-Software und folgt dem Prinzip von 'Infrastructure as Code' – also der Beschreibung von Infrastrukturen durch Code statt durch manuelle Einrichtung. So können andere Organisationen die evaluierte Infrastruktur selbst aufsetzen und an eigene Bedürfnisse anpassen.

Die bisherigen Tests zeigen, dass die Ergebnisse bisher qualitativ hinter denen von Modellen und Infrastruktur großer Anbietern zurückbleibt. Dies zeigt erneut die Problematik der Monopolbildung durch die Kompetenz und Ressourcenkonzentration bei großen Firmen.

Ein Vorteil von selbst gehosteten Modellen liegt jedoch in der umfassenden Kontrolle der Modelle. Unter anderem stellt ParrotPark den Ressourcenverbrauch während der Anwendung dar, was Bewusstsein für den Energieverbrauch bestimmter Aufgaben schaffen kann und so eine fundierte Abwägungen beim Einsatz erleichtert. Außerdem ermöglicht das Hosting eine hohe Kontrolle über eigene Daten. Dies zeigt auch: Es muss nicht immer maximale Performance im Vordergrund stehen. Die Bewertung sollte verschiedene Aspekte wie Kosten, Ressourcenverbrauch, Datenmanagement und mögliche Lock-in-Effekte berücksichtigen.

03 Wie geht's weiter?

Gemeinwohlorientierte Organisationen haben eine entscheidende Verantwortung, Gerechtigkeit bei der Entwicklung und Nutzung von KI zu verankern. Statt reine Profitmaximierung in den Fokus zu stellen, müssen die Bedürfnisse und Perspektiven aller Menschen die Entwicklung und Nutzung von KI gestalten. Durch offene Entwicklungsprozesse können auch die Perspektiven von marginalisierten Gruppen einbezogen werden – insbesondere von jenen, die häufig überproportional von diskriminierenden oder vorurteilsbasierten KI-Systemen betroffen sind. Wenn Organisationen ihre eigenen Ziele unreflektiert über die Erfahrungen von Menschen stellen, die im Arbeitsmarkt Diskriminierung erfahren, als Clickworker:innen ausgebeutet werden oder unter den negativen Umweltfolgen von KI leiden, verstärkt dies Ungleichheiten, statt sie zu bekämpfen. Doch neben der sorgfältigen Abwägung von Anwendungsfällen, Anforderungen und Konsequenzen braucht es auch Experimente und Ausprobieren. Denn das Warten auf perfekte Systeme, die allen Anforderungen entsprechen, kann auch lähmend wirken.

Um unser Handeln dabei zu leiten brauchen wir neben Gerechtigkeit ganz zentral: Solidarität. Dies bedeutet einerseits, mit denjenigen solidarisch zu sein, die am stärksten von bestehenden Ungerechtigkeiten betroffen sind. Andererseits müssen wir Allianzen mit Organisationen schließen, die ähnliche Ziele verfolgen, und dabei bewusst die vielfältigen Stimmen kleinerer Organisationen erhalten. Nur so können wir als Zivilgesellschaft Möglichkeitsräume schaffen und selbstbestimmt die Entwicklung und Nutzung von KI-Anwendungen gestalten.

Mit dem Code of Conduct Demokratische KI setzen wir genau hier an: Gemeinsam mit vielen anderen Organisationen aus der Breite der Zivilgesellschaft entwickeln wir eine Selbstverpflichtung für den Einsatz von KI durch gemeinwohlorientierte Organisationen. Unser bereits veröffentlichtes White Paper zu „KI und Freiheit“¹⁵ sowie ein kommendes White Paper zu „KI und Solidarität“ zeigen auf, wie wir gesellschaftliche Grundwerte verankern können. Damit schaffen wir nicht nur Orientierung für die Praxis, sondern entwickeln auch positive Visionen für einen reflektierten und informierten Umgang mit KI.

¹⁵ Code of Conduct Demokratische KI (2024): Freiheit als zentraler Wert für den Einsatz von KI in der Zivilgesellschaft
https://d-64.org/wp-content/uploads/2024/11/2024_WhitePaper_KI_Freiheit_Demokratische-KI.pdf



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), **Ruven Börger** (ASB Deutschland e.V.), **Esther Finis** (ASB NRW e.V.), **Christoph Hassler** (H:DEI), **Niels Heinemann** (Kompetenzzentrum Technik - Diversity - Chancengleichheit e.V.), **Barbara Jung** (AKTIVOLI-Landesnetzwerk Hamburg e.V.), **Maximilian Kühn** (AWO Bundesverband e.V.), **Leonid Löw** (Deutsche Schachjugend e. V.), **Nevena Nikolajevic** (Correlaid, e.V.), **Damian Paderta** (Offene Kommunen. NRW), **Oliver Pfeiderer** (LAG Selbsthilfe RLP), **Rebecca Roth** (Neue deutsche Medienmacher*innen), **Francesca Schmidt** (netzforma* e.V.), **Jonas Stettner** (Correlaid, e.V.), **Sibylle Würz** (FrauenComputerZentrumBerlin e.V. (FCZB)).

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: D64 / **Johann Lensing**

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

demokratische-ki.de

Mai 2025



Bundesministerium
für Bildung, Familie, Senioren,
Frauen und Jugend