Beschluss Künstliche Intelligenz

Antragsteller*innen:

Antragstext

Positionspapier: Künstliche Intelligenz

Künstliche Intelligenz (KI) ist inzwischen Teil unseres Alltags geworden. Large Language Models (LLMs) [A] wie ChatGPT oder KI-basierte Bild- und Videogeneratoren sind heute für viele Menschen leicht zugänglich und eröffnen neue Möglichkeiten. Gleichzeitig geraten die ökologischen und gesellschaftlichen Folgen von KI-Entwicklung und -Nutzung oft in den Hintergrund.

Ethik und Soziales

Arbeitsmarkt und soziale Ungleichheit

Der zunehmende Einsatz von KI-Systemen [B] wird den Arbeitsmarkt verändern, was vor allem Tätigkeiten mit hohem Anteil digitaler und sprachlich strukturierter Aufgaben von Automatisierung bedroht [1]. Arbeitsplatzverluste oder sinkende Vergütung können dabei soziale Ungleichheiten verstärken. Um dieser Entwicklung entgegenzuwirken, sind angepasste rechtliche Rahmenbedingungen und Arbeitsstrukturen erforderlich [2].

Klickarbeiter*innen und Data-Worker*innen [O] spielen eine zentrale Rolle im Betrieb vieler KI-Systeme, indem sie diese Systeme trainieren, bewerten und mit relevanten Daten versorgen. Ihre Arbeitsbedingungen sind häufig prekär und dur, durch teilweise psychisch belastende Tätigkeiten sowie fehlende soziale Absicherung und unsichere Arbeitsrechte geprägt [3]. Diese Form an global ungleich verteilter digitaler Arbeit kann als ‚digitaler Kolonialismus' [C] bezeichnet werden [4].

Jobs für Berufseinsteiger*innen sind besonders anfällig für Automatisierung durch KI-Systeme. Zwischen 2022 und 2025 konnten deutliche Rückgänge an Beschäftigungen junger Menschen, in besonders betroffenen Bereichen von bis zu 16 %, beobachtet werden [5]. Stimmen in der KI-Industrie halten langfristig den Verlust von bis zu 50 % der Bürojobs realistisch [6].

Datenschutz und Urheberrecht

Die Anwendung der Datenschutz-Grundverordnung (DSGVO) auf KI-Systeme wirft aufgrund umfangreicher und teilweise intransparenter Datenmengen Herausforderungen auf. Insbesondere können die Verarbeitung von personenbezogenen Informationen intransparent sein [7] und die wirksame Wahrnehmung von Betroffenenrechten erschwert werden [8].

Generative KI-Systeme [D] werden teilweise mit urheberrechtlich geschützten Werken trainiert, ohne dass vorher die Zustimmung der Urheber*innen vorliegt. Dies kann insbesondere für freischaffende Künstler*innen problematisch sein, da ihre Werke potenziell Grundlage für konkurrierende Produkte werden [9]. Zudem erschweren täuschend echte KI-generierte Inhalte die Unterscheidung zwischen Original und Reproduktion, was die Marktposition von Künstler*innen schwächt.

Diskriminierung und Inklusion

KI-Systeme können Verzerrungen (Bias) aus ihren Trainingsdaten übernehmen und dadurch diskriminierenden Entscheidungen oder Ergebnisse begünstigen [10]. Besonders betroffen sind marginalisierte Gruppen, die bereits gesellschaftlich-strukturellen Benachteiligungen ausgesetzt sind [11].

Bildung und Kompetenzen

Für einen verantwortungsvollen Umgang mit KI-Systemen ist ein grundlegendes Verständnis ihrer Funktionsweise und Grenzen notwendig. Ein derzeit bestehender Mangel an digitaler und technischer Bildung erschwert jedoch die kritische Bewertung algorithmischer Systeme [E] und kann das Risiko erhöhen, dass Desinformationen nicht erkannt und bestehende soziale Ungleichheiten verstärkt werden [12].

Zudem kann die ständige Verfügbarkeit KI-generierter Antworten eigenständiges Denken, kritische Reflexion und aktive Wissensaneignung beeinträchtigen [13]. Häufig werden KI-Systeme als vermeintlich neutrale und objektive Instanzen wahrgenommen [14]. Dabei bleibt unberücksichtigt, dass diese auf von Menschen ausgewählten und geprägten Daten beruhen [15].

Desinformation und Machtmissbrauch

Generative KI-Systeme wie LLMs können überzeugend wirkende, jedoch inhaltlich fehlerhafte oder erfundene Inhalte erzeugen und so zur Verbreitung von Fehlinformationen [F] beitragen [16]. Werden KI-generierte Inhalte in großem Umfang und ohne ausreichende Kontrolle als Trainingsdaten für nachfolgende Modelle verwendet, kann dies zu einer Verschlechterung der Modellqualität führen und einen Kreislauf aus Informationsverlust und Verzerrung erzeugen (‚model collapse') [17]. Ein ähnliches Phänomen stellt die massenhafte Verbreitung von sogenanntem KI-Slop [G] dar, bei welchem digitale Informationsräume (bspw. soziale Medien [18] [19], Webseiten [20], E-Book-Plattformen [21]) mit massenhaft produzierten, redundanten, qualitativ minderwertigen oder irreführenden KI-generierten Inhalten gefüllt werden [22].

Darüber hinaus können generative KI-Systeme zur Erstellung und Verbreitung von Desinformation [H], Hatespeech [I] und zur (gezielten) Beeinflussung eingesetzt werden [23]. Eine besondere Rolle spielen dabei Deepfakes [J]: Diese Inhalte imitieren reale Personen täuschend echt oder erzeugen realistisch wirkende, aber fiktive Personen, und werden gezielt zur Diskreditierung, für Betrug oder zur Desinformation eingesetzt. Besonders häufig werden sie zur Erstellung nicht einvernehmlicher pornografischer Inhalte missbraucht, wobei überwiegend Frauen betroffen sind [24]. Mit wachsender Qualität und Verbreitung können Deepfakes zunehmend schwer von realen Inhalten zu unterscheiden sein. Die zunehmende Verbreitung von Deepfakes kann damit die Verlässlichkeit des öffentlichen Informationsraums beeinträchtigen und eine Herausforderung für das Vertrauen in demokratische Prozesse und journalistische Medien darstellen [25].

Soziale Folgen

Die zunehmende Nutzung von KI-Systemen kann direkte zwischenmenschliche Interaktionen beeinflussen oder gar ersetzen. Insbesondere KI-gestützte Chatbots können emotionale Bindungen erzeugen, die in Fällen intensiver und problematischer Nutzung zu sozialer Einsamkeit [26] und einer Abnahme von Fähigkeiten der Empathie und Konfliktlösung beitragen [27]. Darüber hinaus

können starke emotionale Abhängigkeiten von KI-basierten Sprachmodellen entstehen [28].

Generative KI-Systeme, vor allem LLMs, tendieren dazu, bestehende Überzeugungen zu bestätigen oder zu verstärken [29] und können damit Filterblasen-Effekte [K] fördern [30]. Kognitive Verzerrungen [L] in den Eingaben von Nutzer*innen, etwa der Bestätigungsfehler [M] oder die Verfügbarkeitsheuristik [N], können die KI-generierten Ergebnisse beeinflussen [31]. Werden Ausgaben zusätzlich personalisiert, steigt die Gefahr selektiver Informationsdarstellung [30]. Diese Prozesse können politische Polarisierung begünstigen [32].

Sicherheitsrisiken durch KI

Führende KI-Forscher*innen schätzen das Risiko für die Auslöschung der Existenz der Menschheit durch KI als realistisch ein [44]. Der Druck, Entwicklungsgeschwindigkeit über die Sicherheit von KI-Systemen zu stellen, ist aktuell hoch. Dies begründet sich in einem unproportional hohen Gewinn bei schneller Entwicklung und führt zu unvorhersehbaren und höchstwahrscheinlich gefährlichen KI-Systemen.
Die Entwicklung der KI funktioniert aktuell nach einem "The Winner takes it all" Prinzip, da der, dessen KI zuerst gewisse Schlüsselfähigkeiten erreicht, einen überproportionalen Gewinn, sei es Geld oder Macht, erzielen kann [44]. Dadurch herrscht ein hoher Druck auf KI-Entwickler Entwicklungsgeschwindigkeit über die Sicherheit der KI-Systeme zu priorisieren. Dies sorgt für eine Weiterentwicklung bei KI welcher unvorhersehbar ist und höchstwahrscheinlich gefährlich.

Klima und Umwelt

Energieverbrauch und Emissionen

Bei der Ressourcennutzung ist zwischen der Trainings- und Nutzungsphase von KI-Systemen zu unterscheiden. Die Verarbeitung einer Textanfrage bei ChatGPT benötigt etwa 2,9 Wh [33], eine Bildgenerierung benötigt je nach Modell zwischen 0,86 bis 4,8 Wh [34]. Eine klassische Google-Suchanfrage im Vergleich benötigt 0,3 Wh [33].

Besonders energieintensiv ist das Training von KI-Systemen, welches je nach Modell mehrere hundert bis über tausend MWh (z. B. 433 mWh für BLOOM, 1.287 MWh für GPT-3) benötigt [35]. Nach Schätzungen benötigten Rechenzentren für KI-Systeme im Jahr 2024 weltweit etwa 415 TWh [36]. Das entspricht ungefähr dem jährlichen Energieverbrauch Deutschlands [37].

Der hohe Energiebedarf führt zudem zu erheblichen Treibhausgasemissionen. In der EU wurden 2023 für jede erzeugte kWh durchschnittlich 207.1 g CO2e produziert, in Deutschland waren es im Schnitt 320 g CO2e [38]. Für das KI-Modell BLOOM wurden beispielsweise während der Trainingsphase 24,7 t CO2e allein durch den Energieverbrauch verursacht. Unter Berücksichtigung aller Prozesse wie der Bauteileproduktion waren es 50,5 t CO2e [39].

Wasserverbrauch

Der Betrieb von Rechenzentren erfordert große Mengen an Wasser zur Kühlung von u. a. Servern [40]. Ein Großteil des Wassers verdunstet im Prozess. Allein das Training von GPT-3 verbrauchte in der Rechenzentrumskühlung 700.000 Liter Wasser. Der komplette Trainingsprozess inklusive der Erzeugung des benötigten

Stroms benötigte 5,4 Millionen Liter Wasser. Auch die Nutzung verursacht indirekt Wasserverbrauch: Eine Anfrage an GPT-3 mit 10 bis 50 (je nach Standort und Kühlung) mittellangen Antworten verbraucht etwa 500 mL Wasser [41].

Energiepolitische Folgen

Die zunehmende Nutzung von KI sorgt für ein erneutes Interesse einiger Unternehmen an Atomenergie, um dadurch die steigenden hohen Energieanforderungen zu decken [42][43]. Die NAJU spricht sich konsequent gegen Atomkraft als Energiequelle aus, da sie weder sicher noch nachhaltig oder generationengerecht ist (vgl. Positionspapier Atomkraft). Grundsätzlich lehnt die NAJU jede Energiequelle ab, die nicht nachhaltig ist, dies gilt neben der Atomkraft ebenso für fossile Energieträger wie Kohle, Öl und Gas (vgl. Positionspapier Keine Braunkohle). Statt auf diese Energieträger zu setzen, muss der zusätzliche Bedarf durch den konsequenten Ausbau erneuerbarer Energien und mehr Effizienz gedeckt werden.

Gesellschaftliche Appelle

Als Jugendumweltverband verstehen wir die verantwortungsvolle und umweltbewusste Nutzung als notwendigen Umgang mit künstlicher Intelligenz in jeglicher Form. Daher appelliert die NAJU an die Gesellschaft, …

… Eigenverantwortung im verantwortungsvollen Umgang mit KI zu zeigen.

… Transparenz in der Datennutzung (z. B. die Verwendung von KI-Tools in der direkten Arbeit mit personenbezogenen Daten öffentlich und deutlich kennzeichnen sowie in der Datenschutzerklärung verankern) zu gewährleisten.

… das Thema "Umgang mit KI" weiter zu tragen sowie das eigene Umfeld für das Thema zu sensibilisieren.

… bestmöglich die Verwendung von KI-generierten Inhalten (z. B. für Printmedien und Social-Media-Kanäle) zu vermeiden.

Unsere Forderungen an die Politik

Aus Sicht der NAJU ist es dringend notwendig, dass die Bundesregierung die Weichen für eine nachhaltige und verantwortungsvolle KI-Nutzung stellt. Wir fordern daher…

1. eine Gesetzgebung, die Anbieter*innen von KI-Systemen verpflichtet, diese Systeme sowie KI-bezogene Infrastruktur langfristig klimaneutral und ressourcenschonend zu betreiben, sowie den Ressourcenverbrauch in der Trainings- und Betriebsphase transparent zu veröffentlichen .

2. Plattformen zu einer stärkeren Kontrolle und Regulierung von KI-generierten Inhalten, insbesondere von Bildern, Videos und Audios, zu verpflichten.

3. einen umfassenden Ausbau von Bildungs- und Aufklärungsangeboten und die Schulung von Fachpersonal zum Thema KI in allen Bereichen.

4. eine konsequente Regulierung und wirksame Kontrolle von KI-generierten Inhalten, die gezielt zur Manipulation von Meinungen oder Verhalten eingesetzt werden.

5. ein Verbot des Einsatzes von KI-generierten Inhalten für politische Werbung.

6. eine Verschärfung und gezielte Weiterentwicklung der Datenschutz-Grundverordnung (DSGVO) im Hinblick auf den Einsatz von KI-Systemen zur Verarbeitung von personenbezogenen Daten.

7. einen besseren rechtlichen Schutz von Künstler*innen im Umgang mit KI-generierten Inhalten, z. B. über Lizenzmodelle oder Beteiligungsrechte, ergänzend zu den bestehenden Opt-out-Regelungen.

8. eine Gesetzgebung, welche die Reproduktion von bestehenden gesellschaftlichen Diskriminierungen durch KI-Systeme verhindert und die Unternehmen hinter KI-Systemen zum aktiven Abbau von diskriminierenden Strukturen verpflichtet.

9. ein Verbot der Generierung und Verbreitung von Inhalten wie Deepfakes, die reale Personen täuschend echt imitieren oder realistisch wirkende Personen darstellen.

10. eine weltweite Verpflichtung zu unabhängigen Sicherheitstests für jedes KI-Modell

Glossar

[A] "Große KI-Sprachmodelle (engl.: Large Language Models – LLMs) sind eine Teilmenge der unimodalen Text-to-Text-Modelle (T2T-Modell), die textuelle Eingaben, sog. Prompts, verarbeiten und darauf basierend Textausgaben erzeugen." (Bundesamt für Sicherheit in der Informationstechnik (2023). Generative KI-Modelle: Chancen und Risiken für Industrie und Behörden . https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative_KI-Modelle.pdf)

[B] "„KI-System“ [ist] ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können;" (Europäisches Parlament und Rat der Europäischen Union (2024). Verordnung (EU) 2024/1689 des europäischen Parlaments und des Rates vom 13. Juni 2024 über Künstliche Intelligenz. http://data.europa.eu/eli/reg/2024/1689/oj)

[C] "Der Begriff „Digitaler Kolonialismus“ beschreibt Macht- und Abhängigkeitsverhältnisse, die durch die globale Digitalisierung entstehen oder verstärkt werden. Technologiekonzerne und Staaten aus den westlichen Industriestaaten – zunehmend auch aus China – kontrollieren digitale Infrastruktur, Plattformen und Datenströme und nutzen Rohstoffe, Arbeitskraft und Daten des sogenannten Globalen Südens unter häufig ausbeuterischen Bedingungen. Konkret zeigt sich dies etwa im Abbau kritischer Rohstoffe für digitale Geräte, in „Clickwork“ oder in der Abschöpfung von Daten ohne angemessene Gegenleistung." (Angela Chukunzira, Nina Grünberger, Philine Janus, Marvin Brodersen für bpb.de (2026). „Das derzeitige Fundament von KI ist nicht tragfähig“ Digitaler Kolonialismus – unsichtbare Arbeits- und Machtverhältnisse hinter KI.

https://www.bpb.de/lernen/digitale-bildung/werkstatt/577276/das-derzeitige-fundament-von-ki-ist-nicht-tragfaehig/)

[D] "Generative KI-Modelle [...] erlernen während des Trainings Muster aus vorhandenen Daten und können in der Folge neue Inhalte wie Texte, Bilder, Audios und Videos erzeugen, die ebenfalls diesen Mustern folgen." (Bundesamt für Sicherheit in der Informationstechnik (2023). Generative KI-Modelle: Chancen und Risiken für Industrie und Behörden. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative_KI-Modelle.pdf)

[E] "Ein Algorithmus ist eine eindeutige Handlungsvorschrift, um ein vorab definiertes Problem zu lösen. [...] Algorithmisches System: Ein System eines oder mehrerer Algorithmen, die in Software implementiert sind, um Daten zu erfassen, zu analysieren und Schlüsse zu ziehen, die zur Lösung eines vorher definierten Problems beitragen sollen." (Innovationsbüro des Bundesministeriums für Familie, Senioren, Frauen und Jugend (2021). Wegweiser "Digitale Debatten", Teil 1: Algorithmische Systeme. https://www.bmbfsfj.bund.de/resource/blob/186298/975f96f2e962ef4354ce9b54c81822c-de/resource/blob/186298/975f96f2e962ef4354ce9b54c81822cde/resource/blob/186298/9-75f96f2e962ef4354ce9b54c81822cde/resource/blob/186298/975f96f2e962ef4354ce9b54c8-1822cde/resource/blob/186298/975f96f2e962ef4354ce9b54c81822cde/resource/blob/186-298/975f96f2e962ef4354ce9b54c81822cde/resource/blob/186298/975f96f2e962ef4354ce9-b54c81822cde/resource/blob/186298/975f96f2e962ef4354ce9b54c81822cde/resource/blo-b/186298/975f96f2e962ef4354ce9b54c81822cde/resource/blob/186298/975f96f2e962ef43-54ce9b54c81822c7/wegweiser-digitale-debatten-teil-1-data.pdf)

[F] "[...] falsche oder irreführende Informationen, die ohne Täuschungsabsicht, also etwa durch mangelnde Recherche, falsche Interpretation von Fakten oder durch Missverständnisse entstehen bzw. verbreitet werden. Dann handelt es sich lediglich um Fehlinformationen [...]." (Bundesamt für Sicherheit in der Informationstechnik (2025). Desinformation im Internet: Eine Bedrohung für Freiheit und Demokratie. https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Onlinekommunikation/Soziale-Netzwerke/Sichere-Verwendung/Desinformation/desinformation_node.html)

[G] "Als Slop (auch KI-Slop, [...]) werden Medieninhalte – einschließlich Text und Bilder – von niedriger inhaltlicher Qualität bezeichnet, die mit Hilfe von generativer künstlicher Intelligenz erzeugt werden." (Wikipedia – Die freie Enzyklopädie. (7. Oktober 2025). "Slop". https://de.wikipedia.org/w/index.php?title=Slop&oldid=263763778)

[H] "Desinformation ist die Verbreitung falscher oder irreführender Informationen, um Einzelpersonen, Gruppen oder die öffentliche Meinung als Ganzes zu beeinflussen. Eine Desinformation liegt vor, wenn sie nach objektiven Maßstäben inhaltlich unzutreffend ist, der Urheber dies weiß und sie dennoch mit dem Ziel der Beeinflussung verwendet. Gleiches gilt für das Verschweigen wesentlicher Teile einer Information." (Bundesamt für Verfassungsschutz (2025). https://www.verfassungsschutz.de/SharedDocs/glossareintraege/DE/D/desinformation-de/SharedDocs/glossareintraege/DE/D/desinformationde/SharedDocs/glossareintraege-/DE/D/desinformationde/SharedDocs/glossareintraege/DE/D/desinformationde/SharedD-ocs/glossareintraege/DE/D/desinformationde/SharedDocs/glossareintraege/DE/D/desi-nformationde/SharedDocs/glossareintraege/DE/D/desinformationde/SharedDocs/glossa-reintraege/DE/D/desinformationde/SharedDocs/glossareintraege/DE/D/desinformation-de/SharedDocs/glossareintraege/DE/D/desinformation.html)

[I] "Hatespeech ist ein Oberbegriff für verbal oder schriftlich geäußerte, gruppenbezogene Menschenfeindlichkeit. Dazu zählen u. a. Rassismus, Sexismus oder Antisemitismus. Unter Hatespeech fallen Beleidigungen, Aufrufe zur Gewalt, Drohungen und weitere Äußerungen, unabhängig davon, ob diese strafbar sind oder nicht. Der Begriff Hatespeech ist allerdings nicht ganz klar definiert und wird manchmal auch dann verwendet, wenn ganz allgemein Hass im Internet gemeint ist." (HateAid (2023). Hatespeech: Von Sprache als Gefahr. https://hateaid.org/hatespeech/)

[J] "Deepfakes sind aufwändig erstellte beziehungsweise verfälschte Bilder, Tonaufzeichnungen oder Videos. Sie entstehen auch unter Einsatz Künstlicher Intelligenz (KI) und dienen insbesondere der Desinformation." (Bundesamt für Verfassungsschutz (2025). https://www.verfassungsschutz.de/SharedDocs/glossareintraege/DE/D/deepfake.html)

[K] "Als Filterblase, engl. »Filter Bubble« [Pariser 2012], wird der Effekt bezeichnet, im Internet bevorzugt Suchergebnisse zu erhalten, die aufgrund persönlicher Neigungen zum eigenen Welt bild passen." (Kompetenzzentrum Öffentliche IT (2016). Digitalisierung des Öffentlichen. https://www.oeffentliche-it.de/publikationen/digitalisierung-des-oeffentlichen/Digitalisierung des Öffentlichen.pdf)

[L] "Kognitive Verzerrungen oder kognitive Fehler sind ein kognitionspsychologischer Sammelbegriff: Eigenschaften der zu verarbeitenden Information werden dabei systematisch verzerrt verstanden bzw. mental repräsentiert. […] Dies kann ggf. fehlerhafte oder suboptimale Entscheidungen oder Handlungen forcieren." (Wikipedia – Die freie Enzyklopädie. (23. November 2025). "Kognitive Verzerrung". https://de.wikipedia.org/w/index.php?title=Kognitive_Verzerrung&oldid=261799928)

[M] "Ein Bestätigungsfehler (auch Bestätigungsfehlschluss oder Bestätigungsverzerrung, engl. confirmation bias) ist ein Begriff der Kognitionspsychologie, der die Neigung bezeichnet, Informationen so zu ermitteln, auszuwählen und zu interpretieren, dass diese die eigenen Erwartungen erfüllen (bestätigen)." (Wikipedia – Die freie Enzyklopädie. (2. Februar 2026). "Bestätigungsfehler" . https://de.wikipedia.org/w/index.php?title=Bestätigungsfehler&oldid=263930666)

[N] "Verfügbarkeitsheuristik (englisch availability heuristic) ist eine verkürzende kognitive Operation, die zu Urteilsfehlern führt. Sie ersetzt die schwierige Frage nach der Häufigkeit eines Ereignisses oder dem Umfang einer Kategorie durch die einfachere Frage, wie leicht es fällt, sich an passende Beispiele zu erinnern." (Wikipedia – Die freie Enzyklopädie. (14. Juli 2025). "Verfügbarkeitsheuristik". https://de.wikipedia.org/w/index.php?title=Verfügbarkeitsheuristik&oldid=2578878-71php?title=Verfügbarkeitsheuristik&oldid=257887871php?title=Verfügbarkeitsheuristik&oldid=257887871php?title=Verfügbarkeitsheuristik&oldid=257887871php?title=Verfügbarkeitsheuristik&oldid=257887871php?title=Verfügbarkeitsheuristik&oldid=257887871php?title=Verfügbarkeitsheuristik&oldid=257887871php?title=Verfügbarkeitsheuristik&oldid=257887871)

[O] "Klickarbeiter*innen und Data-Worker*innen" sind Menschen, die KI-Systeme

weltweit mit Daten versorgen. Sie bennen zum Beispiel Gegenstände auf Fotos, klassifizieren Texte oder transkribieren Tonaufnahmen, sodass diese Informationen als Trainingsmaterial für KI-Modelle genutzt werden können. Ihre Arbeit ist für die Entwicklung von KI essentiell, bleibt aber weitgehend unsichtbar und findet häufig unter prekären Bedingungen statt. (Der Spiegel 2026 - "So sieht die Arbeit von Clickworkern aus". https://www.spiegel.de/netzwelt/web/training-fuer-kuenstliche-intelligenz-so-sieht-die-arbeit-von-clickworkern-aus-a-7c018325-440d-461c-a004-6b3b35fb9ed1"

Literaturverzeichnis

[1] Gmyrek, P., Berg, J., Kamiński, K., Konopczyński, F., Ładna, A., Nafradi, B., Rosłaniec, K., & Troszyński, M. (2025). Generative AI and jobs: a refined global index of occupational exposure. https://doi.org/10.54394/HETP0387

[2] Lane, M. (2024). Who will be the workers most affected by AI?: A closer look at the impact of AI on women, low-skilled workers and other groups. https://doi.org/10.1787/14dc6f89-en

[3] Deutschlandfunk (2024). Data Worker - Die harte Arbeit der menschlichen KI-Trainer. https://www.deutschlandfunk.de/ki-data-worker-clickworker-deep-learning-100.html

[4] Angela Chukunzira, Nina Grünberger, Philine Janus, Marvin Brodersen für bpb.de (2026). „Das derzeitige Fundament von KI ist nicht tragfähig“ Digitaler Kolonialismus – unsichtbare Arbeits- und Machtverhältnisse hinter KI.

https://www.bpb.de/lernen/digitale-bildung/werkstatt/577276/das-derzeitige-fundament-von-ki-ist-nicht-tragfaehig/

[5] Brynjolfsson, E., Chandar, B., Chen, R. (2025). Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. https://digitaleconomy.stanford.edu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Noedu/app/uploads/2025/11/CanariesintheCoalMine_Nov25.pdf

[6] Morris, C. (2025). Anthropic CEO warns AI could eliminate half of all entry-level white-collar jobs. https://fortune.com/2025/05/28/anthropic-ceo-warning-ai-job-loss/

[7] Lucchi, N., Hunter, S. (2025). Generative AI and Copyright: Training, Creation, Regulation. https://www.europarl.europa.eu/thinktank/en/document/IUST_STU(2025)7-eu/thinktank/en/document/IUST_STU(2025)7eu/thinktank/en/document/IUST_STU(2025)7-eu/thinktank/en/document/IUST_STU(2025)7eu/thinktank/en/document/IUST_STU(2025)7-eu/thinktank/en/document/IUST_STU(2025)7eu/thinktank/en/document/IUST_STU(2025)7-eu/thinktank/en/document/IUST_STU(2025)774095

[8] Pasetti, M., Santos, J.W., Corrêa, N.K. et al. (2025). Technical, legal, and ethical challenges of generative artificial intelligence:

an analysis of the governance of training data and copyrights. https://doi.org/10org/10org/10org/10org/10org/10org/10org/10org/10.1007/s44163-025-00379-6

[9] Hutson, J. (2024). The Evolving Role of Copyright Law in the Age of AI-Generated Works. https://doi.org/10.21202/jdtl.2024.43

[10] Deckker, D., & Sumanasekara, S. (2025). Bias in AI Models: Origins, Impact, and Mitigation Strategies. https://doi.org/10.20944/preprints202503.1629.v1

[11] Sun et al. (2025). Uncovering Bias in Foundation Models: Impact, Testing, Harm, and Mitigation. https://arxiv.org/abs/2501.10453

[12] European Commission (2025). The State of the Digital Decade 2025 report. https://digital-strategy.ec.europa.eu/en/policies/2025-state-digital-decade-package

[13] Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. https://www.mdpi.co-cococococococococom/2075-4698/15/1/6

[14] Gerlich, M. (2024). Exploring Motivators for Trust in the Dichotomy of Human—AI Trust Dynamics. https://doi.org/10.3390/socsci13050251

[15] Wan, He et al. (2023). BiasAsker: Measuring the Bias in Conversational AI System. https://doi.org/10.1145/3611643.3616310

[16] Park, S., Nan, X. (2025). Generative AI and misinformation: a scoping review of the role of generative AI in the generation, detection, mitigation, and impact of misinformation. https://doi.org/10.1007/s00146-025-02620-3

[17] Shumailov, I., Shumaylov, Z., Zhao, Y. et al. (2024). AI models collapse when trained on recursively generated data. https://doi.org/10org/10org/10org/10org/10org/10org/10org/10org/10.1038/s41586-024-07566-y

[18] Sun, Z. et al. (2024). Are We in the AI-Generated Text World Already? Quantifying and Monitoring AIGT on Social Media. https://arxiv.org/abs/2412.18148

[19] Gillham J. (2024). Over ½ of Long Posts on LinkedIn are Likely AI-Generated Since ChatGPT Launched. https://originality.ai/blog/ai-content-published-linkedin

[20] Law, R., Guan, X., Soulo, T. (2025). 74% of New Webpages Include AI Content (Study of 900k Pages). https://ahrefs.com/blog/what-percentage-of-new-content-is-ai-generated/

[21] Brodsky S. (2024). The AI-generated Books Trend is Getting Worse. https://aibusiness.com/responsible-ai/the-ai-generated-books-phenomenon-is-getting-worse

[22] Stanusch, N. et al. (2025). AI-Generated Algorithmic Virality.

https://arxiv.org/abs/2508.01042

[23] Park, S., Nan, X. (2025). Generative AI and misinformation: a scoping review of the role of generative AI

in the generation, detection, mitigation, and impact of misinformation. https://doi.org/10.1007/s00146-025-02620-3

[24] Lazard, L., Capdevila, R., Turley, E. L., Gilfoyle, K., & Stavropoulou, N. (2025). Deepfake Technology and Gender-Based Violence: A scoping review. https://doi.org/10.1177/15248380251384271

[25] Barari, S., Lucas, C., & Munger, K. (2024). Political deepfakes are as credible as other fake media and (Sometimes) real media. https://doi.org/10.1086/732990

[26] Folk, D., & Dunn, E. (2026). How Does Turning to AI for Companionship Predict Loneliness and Vice Versa?. https://doi.org/10.1177/09567976261427747

[27] Talmaţchi, M.L. (2025). The Influence of Artificial Intelligence on Communication. https://doi.org/10.2478/bsaft-2025-0014

[28] Chu, M.D., Gerard, P., Pawar, K., Bickham, C., & Lerman, K. (2025). Illusions of Intimacy: Emotional Attachment and Emerging Psychological Risks in Human-AI Relationships. https://doi.org/10.48550/arXiv.2505.11649

[29] Rrv, A., Tyagi, N., Uddin, M. N., Varshney, N., & Baral, C. (2024). Chaos with Keywords: Exposing Large Language Models Sycophancy to Misleading Keywords and Evaluating Defense Strategies. https://doi.org/10.18653/v1/2024.findings-acl.755

[30] Lazovich, T. (2023). Filter bubbles and affective polarization in user-personalized large language model outputs. https://proceedings.mlr.press/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lapress/v239/lazovich23a.html

[31] Sun, Y., & Kok, S. (2025). Investigating the Effects of Cognitive Biases in Prompts on Large Language Model Outputs. https://arxiv.org/abs/2506.12338

[32] Rodilosso, E. (2024). Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization. https://doi.org/10.1007/s13347-024-00758-4

[33] International Energy Agency: Electricity 2024. https://iea.blob.core.windows.net/assets/6b2fd954-2017-408e-bf08-952fdd62118a/Electricity2024-Analysisandforecastto2026.pdf

[34] Bertazzini, G., Albisani, C., Baracchi, D., et al. (2025). The Hidden Cost of an Image: Quantifying the Energy Consumption of AI Image Generation. https://arxiv.org/html/2506.17016v1

[35] de Vries, A. (2023). The growing energy footprint of artificial intelligence. https://doi.org/10.1016-/j1016/j1016/j1016/j1016/j1016/j1016/j1016/j1016/j1016/j.joule.2023.09.004

[36] IEA (2025). Energy and AI. https://www.iea.org/reports/energy-and-ai

[37] Bundesnetzagentur: Der Strommarkt im Jahr 2024. https://www.smard.de/page/home/topic-article/211784/215556

[38] European Environment Agency. (2025). Greenhouse gas emission intensity of electricity generation in Europe.

https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emission-intensity-of-1

[39] Luccioni, A., Viguier, S., Ligozat, A. (2023). Journal of Machine Learning Research. https://www.jmlr.org/papers/v24/23-0069.html

[40] Jennings, C. (2025). The Cloud is Drying our Rivers: Water Usage of AI Data Centers. https://ethicalgeo.org/the-cloud-is-drying-our-rivers-water-usage-of-ai-data-centers/

[41] Li, P., Yang, J., Islam, M., et al. (2025). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. https://arxiv.org/abs/2304.03271

[42] da Silva, J. (2024). Google turns to nuclear to power AI data centres. https://www.bbc.com/news/articles/c748gn94k95o

[43] Browne, R. (2024). Sustainable Future: Why Big Tech is turning to nuclear to power its energy-intensive AI ambitions. https://www.cnbc.com/2024/10/15/big-tech-turns-to-nuclear-energy-to-fuel-power-intensive-ai-ambitions.html

[44] Markov, G., Segerie, C.-R. (2025). The AI Risk Spectrum: From Dangerous Capabilities to Existential Threaths. https://arxiv.org/pdf/2508.13700