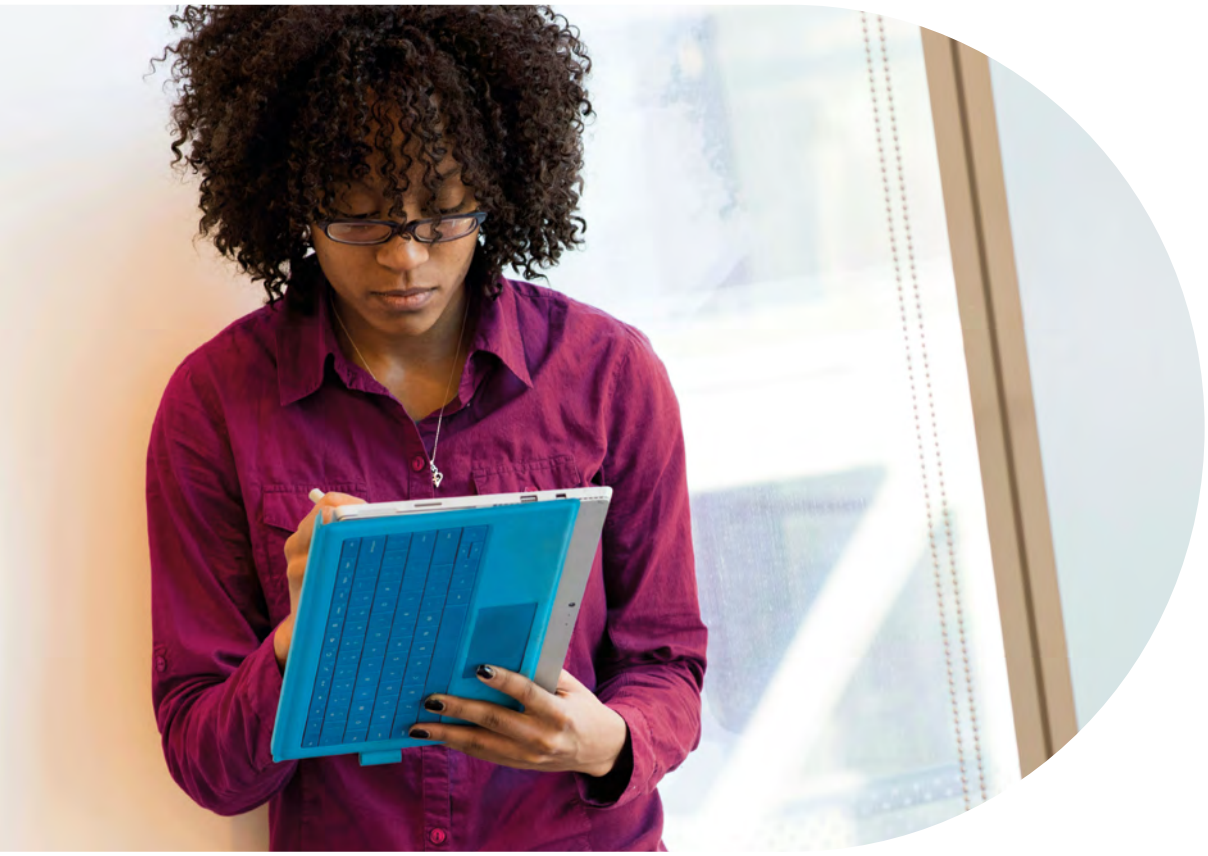
AdB Arbeitskreis deutscher Bildungsstätten e. V.



Künstliche Intelligenz verstehen und nutzen

Arbeitshilfe zum Einsatz von KI
in Organisationen und Bildungsarbeit

Arbeitskreis deutscher Bildungsstätten e. V. (AdB)

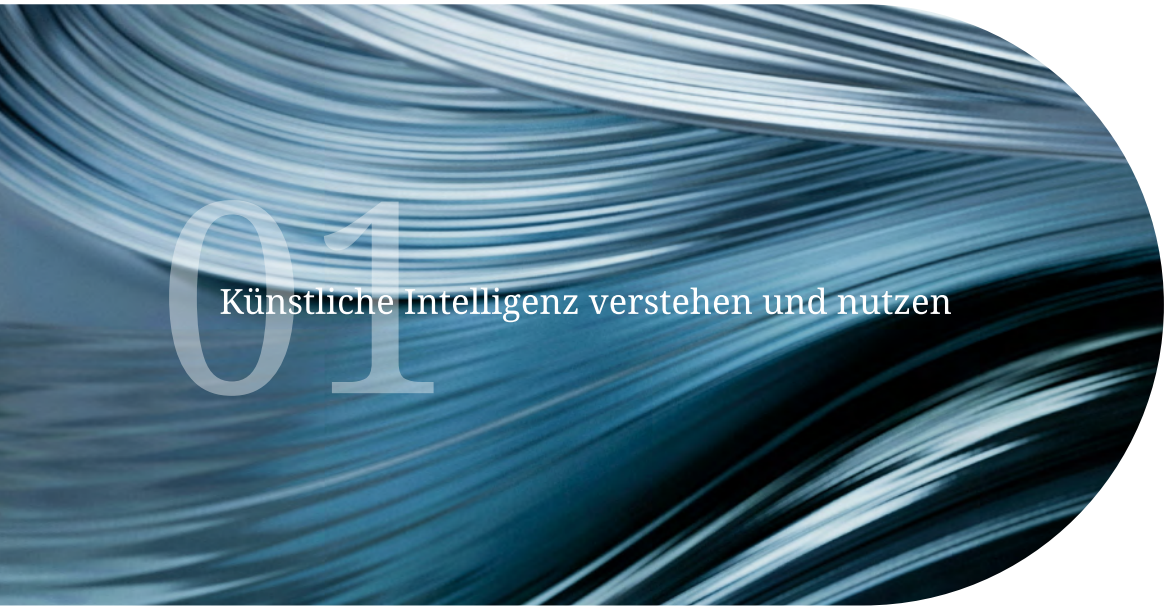
Arbeitskreis deutscher Bildungsstätten e.V. (AdB)
Mühlendamm 3, 10178 Berlin

(030) 400 401 00
info@adb.de
adb.de

Der AdB wird gefördert aus Mitteln des Kinder- und Jugendplans des Bundes.

Bundesministerium
für Familie, Senioren, Frauen
und Jugend

Inhalt



01

Künstliche Intelligenz verstehen und nutzen

KI – nicht richtig intelligent, aber mitunter nützlich

Spätestens seit 2023 sind Anwendungen Künstlicher Intelligenz in der breiten Öffentlichkeit angekommen. Mit Diensten wie ChatGPT, DALL-E oder Midjourney waren Tools mit bisher ungeahnten Fähigkeiten plötzlich für Endanwender*innen einfach verfügbar.

Im täglichen Umgang haben sich KI-Systeme vor allem in Bereichen ausgebreitet, die bisher als besondere menschliche Kompetenz galten: in Kreativprozessen und beim Umgang mit Sprache. Derweil haben die Large Language Models, auf den KI derzeit basiert, mit „Intelligenz“ eigentlich nichts zu tun.

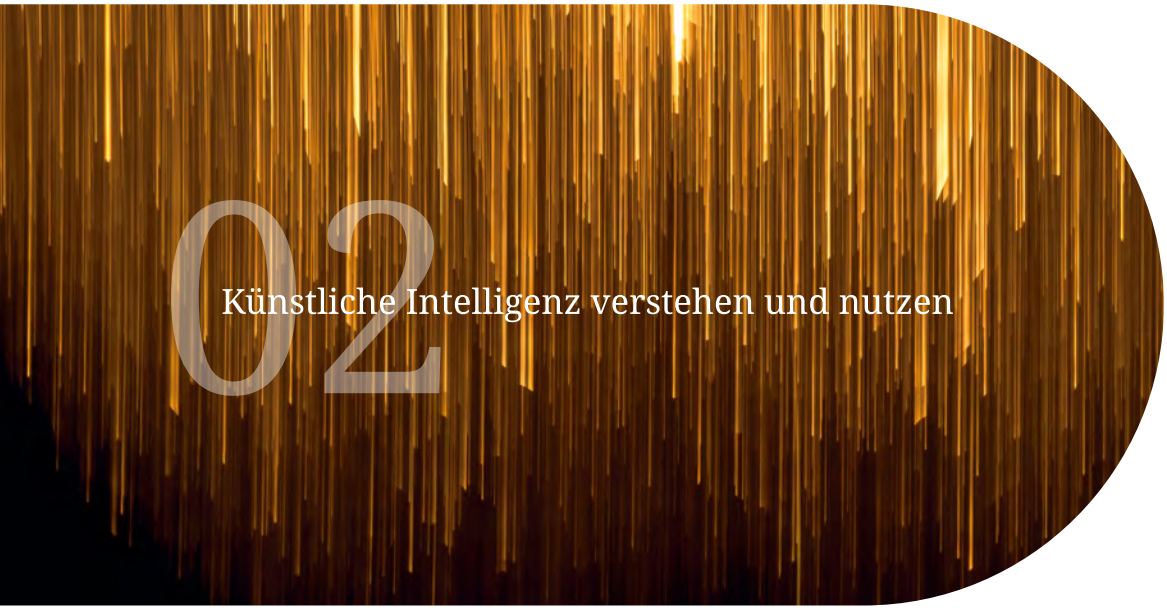
Trainiert mit unermesslich großen Datenbeständen errechnen neuronale Netze ihre Antworten auf Basis von Wahrscheinlichkeiten. Logik und echtes Verständnis sind nicht vorhanden.

Dennoch zeigen sich die neuen Systeme in vielerlei Hinsicht als hilfreich: Texte können ausformuliert, zusammengefasst, ergänzt, auf Syntax und Stil geprüft oder natürlichsprachlich übersetzt werden. Bilder und ganze Videosequenzen werden auf Basis von Text-Prompts generiert oder verändert. Chatbots können mehrheitlich sinnvoll erscheinende Unterhaltungen führen, Fragen beantworten und als teil-autonome Agenten ganze Workflows bearbeiten. Töne und Musik werden nach Bedarf generiert, sprachliche Eingaben transkribiert. Die bisherigen Grenzen zwischen Video, Audio und Text werden fluide.

„Die neuen Möglichkeiten, die sich durch generative KI auftun, begeistern schnell, verursachen aber auch viele Fragezeichen."

Die neuen Möglichkeiten, die sich durch generative KI auftun, begeistern schnell, verursachen aber auch viele „Fragezeichen". Unsicherheiten bezüglich Vertraulichkeit, Verlässlichkeit und Objektivität kommen auf. Computer-generierte Profilbildung und unkontrollierbare, persönlich maßgeschneiderte Ansprache durch KI-Systeme bergen kaum überschaubare Risiken. Bilderkennende und -interpretierende Verfahren werden zunehmend verzahnt mit Sensorik und Robotik. Während diese einerseits die medizinische Analyse unterstützen oder helfen kann, Straftäter*innen zu überführen, ist andererseits der Weg zu vorausschauender Polizeiarbeit und möglichen Vorverurteilungen nicht weit. Sicher diskriminierungsfrei sind menschengemachte Algorithmen allemal nicht. Gesellschaftliche Folgen könnten gewaltig sein.

Diese Arbeitshilfe möchte Akteur*innen in der politischen Bildung dabei unterstützen, den Einsatz von Systemen künstlicher Intelligenz bewusst(er) anzugehen. In den folgenden Kapiteln stellen wir den rechtlichen Rahmen vor und beschäftigen uns etwas tiefer mit Fragen von Ethik, Urheberrecht, Deepfakes und unbeabsichtigten Datenabflüssen. Die zweite Hälfte der Arbeitshilfe beinhaltet eine detaillierte Checkliste für den Einsatz von KI bei Trägern und Einrichtungen.



In Arbeit: Der rechtliche Rahmen

Bahnbrechende Entwicklungen benötigten Zeit, um in der Gesellschaft anzukommen und bis sich die Tragweite der Veränderungen abzeichnet. Beim Thema Künstlicher Intelligenz war die Europäische Union schon frühzeitig aktiv.

Dies gilt umso mehr, wenn man bedenkt, dass die Idee der „Transformer“, die den aktuellen Innovationssprung bei KI-Systemen möglich gemacht haben, erst 2017 entstand (vgl. Wikipedia 2024; Ernst 2024). Bereits im Jahr 2018 veröffentlichte die EU-Kommission ein Strategiepapier „Künstliche Intelligenz für Europa“ (Europäische Kommission 2018). Der dazugehörige „Koordinierte Plan für künstliche Intelligenz“ wurde mehrfach ergänzt (Europäische Kommission 2021). 2020 folgte dann das Weißbuch zu Künstlicher Intelligenz (Europäische Kommission 2020).

Diese Vorarbeiten fanden Eingang in das europäische „Gesetz über künstliche Intelligenz“ (kurz: KI-Verordnung/AI Act), über das im Dezember 2023 nach langen Verhandlungen eine Einigung erzielt wurde und das am 21. Mai 2024 vom Rat der 27 EU-Mitgliedstaaten verabschiedet werden konnte (vgl. Bundesregierung 2024). Das Regelwerk ist als EU-Verordnung angelegt und gilt daher ähnlich wie die Datenschutz-Grundverordnung in allen EU-Mitgliedsstaaten (im Unterschied zu EU-Richtlinien, die erst in nationale Gesetze gegossen werden müssen).

Die KI-Verordnung trat am 1. August 2024 in Kraft und gilt für alle Menschen, die in der EU leben, dort KI-Systeme anbieten oder einsetzen. Anders als die DSGVO gilt die KI-Verordnung auch für staatliche Stellen. Hierzu haben sich die Mitgliedsstaaten allerdings einen Vorbehalt eingeräumt: Wenn es um Belange „nationaler Sicherheit“ und um militärische Zwecke geht, soll die KI-Verordnung keine Anwendung finden.

Der Leitgedanke der Verordnung ist, dass KI keine negativen Auswirkungen auf Sicherheit, Gesundheit und Grundrechte haben soll. Das Individuum steht im Zentrum dieses Schutzgedankens. In der Ausgestaltung verfolgt die KI-Verordnung daher einen risikobasierten Ansatz: Risikoreiche Systeme sind verboten, Hochrisiko-Systeme sind reglementiert und risikoarme Systeme haben keine Auflagen. Auch Bildungsanwendungen können dabei unter die Hochrisiko-Systeme fallen – z. B. dann, wenn sie eine Profilbildung über konkrete Nutzer*innen zulassen. Anbieter von Hochrisiko-Systemen unterliegen einer Registrierungspflicht. Eine öffentlich einsehbare Datenbank dieser Systeme soll etabliert werden. Allerdings basieren die Einträge erst einmal auf einer Selbstauskunft der jeweiligen Hersteller.

Verboten sind u. a. Social-Scoring-Systeme (eine Art „Bepunktung der Vertrauenswürdigkeit“ einer Person, wie sie u. a. in China etabliert wurde) oder

i

Zum Weiterlesen

Die Rechtsvorschriften zur künstlichen Intelligenz der EU

Der AI Act Explorer:

Die Datenschutz-Grundverordnung der EU mit Erwägungsgründen

Die DSGVO:

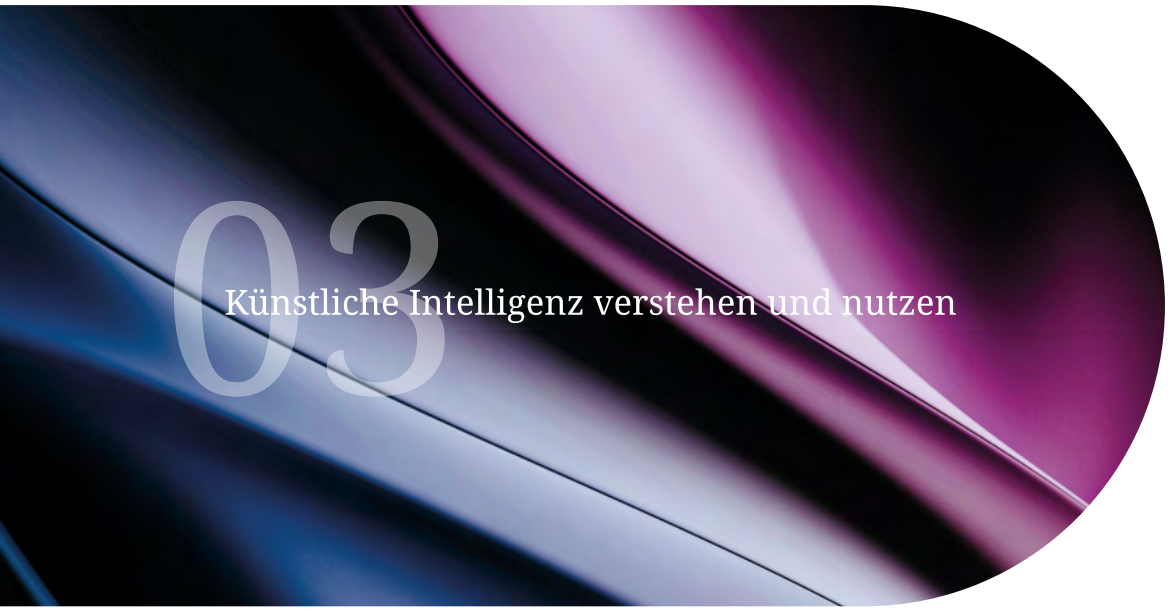
die Emotionserkennung am Arbeitsplatz. Weiterhin sind digitale Wasserzeichen für KI-generierte Medien vorgesehen, um Desinformation zu verhindern.

Die KI-Verordnung führt neue Informationsrechte für Betroffene ein: Neu sind sowohl das Recht der Information, ob man einem Hochrisiko-System ausgesetzt wird als auch das Recht zu erfahren, warum das jeweilige System eingesetzt wird. Anders als bei der DSGVO sind zwar Bußgelder für Verstöße, aber kein Schadenersatz für Betroffene vorgesehen. Ab August 2025 müssen Anbieter von „General Purpose AI", also z. B. den großen Large Language Models wie ChatGPT, ihre Trainingsdaten offenlegen (vgl. EU Artificial Intelligence Act 2024).

Zwischen den beratenden EU-Institutionen sehr umstritten waren Themen einer vorausschauenden Polizeiarbeit und die biometrische Gesichtserkennung im öffentlichen Raum. Hier haben die Mitgliedsstaaten durch den Vorbehalt nationaler Sicherheit dennoch weitreichende Möglichkeiten erhalten.

Die Europäische Union preist die KI-Verordnung als weltweit erste KI-Regulierung an. Das trifft aber nur eingeschränkt zu:

- In den USA wurden im Oktober 2023 mittels eines Dekrets des Präsidenten erste Richtlinien zu Künstlicher Intelligenz herausgegeben (vgl. The White House 2023). Diese „Executive Order" basiert auf einer Verabredung mit 15 Technologiefirmen. Hersteller müssen vor Veröffentlichung eines KI-Systems Testungen durchführen und diese Ergebnisse den Behörden mitteilen, sobald es um Gefahren für nationale Sicherheit, die Wirtschaft oder die öffentliche Gesundheit geht. Zudem wird die Entwicklung von Standards zur Markierung und Erkennung von KI-generierten Inhalten ins Auge gefasst. Der Schutz von Privatsphäre findet ebenfalls Erwähnung.
- Zur weiteren Ausgestaltung des rechtlichen Rahmens verabschiedete der Gesetzgeber im Bundesstaat Kalifornien – also auch gültig für das Silicon Valley – Ende August 2024 einen „AI Act", der ähnlich umfassend wie die europäische KI-Verordnung ausfällt (vgl. California Legislative Information 2024). Durch das Veto des Gouverneurs von Kalifornien treten davon jedoch nur Teile in Kraft: Während Deepfakes reguliert werden, sollen Firmen in ihren Tätigkeiten kaum eingeschränkt werden.
- In China wurden bereits mehrere Gesetze zu unterschiedlichen Aspekten von Künstlicher Intelligenz erlassen. Sie untersagen unter anderem automatische Entscheidungsfindung, versuchen die Ergebnisverzerrung (Bias) von KI-Systemen zu verbieten und schreiben seit 2022 vor, dass KI-generierte Inhalte entweder wahrheitsgetreu sein oder die Zustimmung Betroffener haben müssen (z. B. bei Deepfakes). Diese Regelungen gelten allerdings nur für Unternehmen und Privatpersonen, nicht aber für staatliche Stellen/KI-Systeme (vgl. Hutson 2023). Offizielles Ziel ist der Ausgleich zwischen Innovation und Sicherheit.



Wem „gehören“ KI-generierte Inhalte?

Häufig werden mit Hilfe von KI Inhalte dazu generiert, um sie in Online- oder Offline-Veröffentlichungen zu verwenden. Aber kann man das so einfach?

Die rechtliche Situation in Deutschland ist auf den ersten Blick eindeutig: Im § 2 Abs. 2 des Urheberrechtsgesetzes steht: „Werke im Sinne dieses Gesetzes sind nur persönliche geistige Schöpfungen.“ (Bundesministerium der Justiz/Bundesamt für Justiz 2024 a)

Mit anderen Worten: Damit ein Werk nach dem Urheberrecht geschützt sein kann, muss es das Werk eines *Menschen* sein, dem zudem eine gewisse *Schöpfungshöhe* innewohnt. Da KI-Inhalte nicht von einem Menschen erstellt sind und KI-Prompts häufig eher schlichte Anweisungen ohne eben diese Schöpfungshöhe beinhalten, sind nach dem Urheberrecht die meisten KI-generierten Werke nicht geschützt.

Erst wenn ein hoher Mitwirkungsgrad eines Menschen nachzuweisen ist, können KI-generierte Werke einen Schutz nach dem Urheberrecht erhalten. Das wäre z. B. der Fall, wenn man ein großes Gesamtwerk erstellt, bei dem KI-Tools nur einzelne Schritte unterstützen. Auch können „gute" Prompts möglicherweise Schutzrecht genießen.

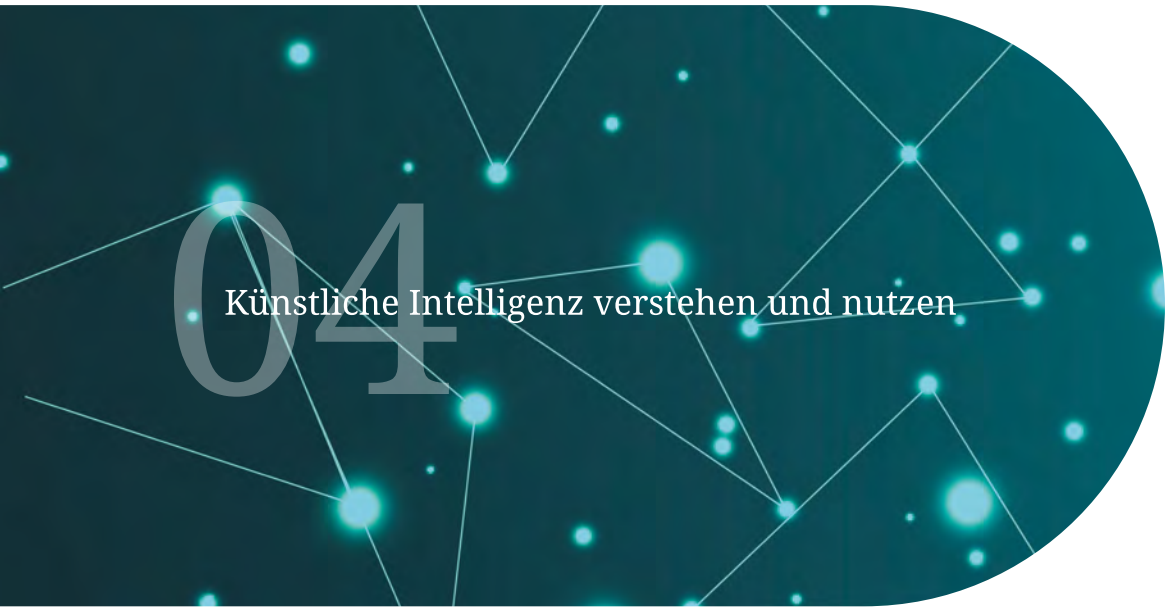
Diese Interpretation ist derzeit in der EU und in den USA üblich. Aber bereits in Großbritannien und Indien können computergenerierte Werke hingegen direkt geschützt werden. Das verkompliziert den Umgang erheblich: Schutzrechte können sowohl die Software-Ersteller*innen des KI-Tools und auch die KI-Anwender*innen in Anspruch nehmen.

„Nicht übersehen werden darf dabei, dass KI-generierte Werke Persönlichkeitsrechte oder Markenrechte Dritter beachten müssen."

Nicht übersehen werden darf dabei, dass KI-generierte Werke Persönlichkeitsrechte oder Markenrechte Dritter beachten müssen: Bilder, die ganz offensichtlich eine existierende Person oder bekannte Figuren wie z. B. die Simpsons darstellen, dürfen also ebenso nicht einfach veröffentlicht werden.

Weiterhin beeinflusst die Frage nach dem Ursprung der Trainingsdaten in KI-Systemen die Diskussion um das Urheberrecht. Autor*innen, Verwertungsgesellschaften und Medienkonzerne führen in letzter Zeit vermehrt juristische Verfahren mit KI-Herstellern, um zu klären inwieweit KI-Training Lizenzen benötigt oder z. B. unter das amerikanische Prinzip des „Fair Use" fällt. Es ist davon auszugehen, dass diese rechtlichen Auseinandersetzungen noch einige Jahre andauern werden.

In Deutschland ist nach § 60d des UrhG „Text und Data Mining für Zwecke der wissenschaftlichen Forschung" generell zulässig (Bundesministerium der Justiz/Bundesamt für Justiz 2024b).



Deepfakes und Persönlichkeitsrechte

Mit Hilfe von KI-Werkzeugen ist es so einfach wie nie, audiovisuelle Inhalte zu erstellen: Generierte „Fotos“, die Nachahmung bekannter Stimmen oder sogar ganze Videosequenzen sind auch mit geschultem Blick/Ohr immer schwerer von „echten“ Darstellungen zu unterscheiden. Diese sogenannten Deepfakes bringen zunehmend die Verlässlichkeit und damit das Vertrauen in Informationsquellen ins Wanken.

Mit Hilfe von KI-Werkzeugen ist es so einfach wie nie, audiovisuelle Inhalte zu erstellen: Generierte „Fotos“, die Nachahmung bekannter Stimmen oder sogar ganze Videosequenzen sind auch mit geschultem Blick/Ohr immer schwerer von „echten“ Darstellungen zu unterscheiden. Diese sogenannten Deepfakes bringen zunehmend die Verlässlichkeit und damit das Vertrauen in Informationsquellen ins Wanken.

Für die Erstellung und Verwendung von Deepfakes gibt es in Deutschland bisher keine speziellen rechtlichen Regelungen (vgl. Deutscher Bundestag/Wissenschaftliche Dienste 2024). Es gelten die bestehenden Persönlichkeits- und Grundrechte. Eine mögliche Handhabe gegen absichtlich irreführende Deepfakes könnte das Recht auf Informationsfreiheit (Art. 5 Abs. 1 Grundgesetz) darstellen.

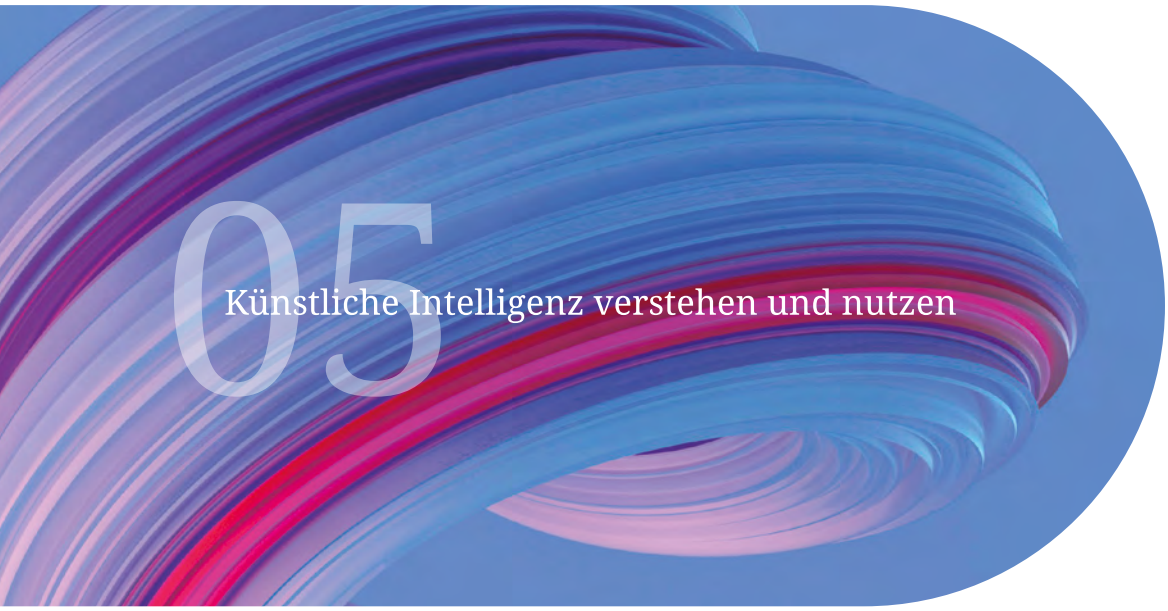
Der „Digital Services Act" (zu Deutsch „Gesetz über digitale Dienste"), der seit 17. Februar 2024 europaweit gilt, schreibt Betreibern sehr großer Online-Plattformen jedoch eine Kennzeichnung von Deepfakes vor, falls man jene für echt halten könnte.[1] Auf EU-Ebene wird darüber beraten, ob die Kennzeichnungspflicht auf alle künstlich hergestellten Werke erweitert werden soll.

In vielen Fällen dienen Deepfakes der Unterhaltung oder Bildung. In solchen Kontexten sind sie häufig auch ohne Kennzeichnung zu erkennen. Problematisch wird es, wenn Deepfakes dazu genutzt werden, die Betrachtenden absichtlich zu täuschen oder abgebildete Personen zu schädigen. Einsatzgebiete dieser Täuschung umfassen z. B. Desinformationskampagnen, Verleumdung, Überwindung biometrischer Verifizierungssysteme und das Abgreifen von vertraulichen Informationen oder Geldmitteln via Social Engineering (z. B. als modernisierte Variante des „Enkeltricks").

Eine weitere, für Betroffene sehr unangenehme Situation, ergibt sich aus künstlich generierten Nacktbildern oder pornografischen Werken. Die Täter sind fast immer männlich. Häufig stammen sie aus dem privaten Umfeld oder handeln voyeuristisch oder politisch motiviert (vgl. Westarp 2022). Für die Betroffenen – fast ausschließlich Frauen – ist es sehr aufwändig und extrem schwierig, Bildmaterial, das im Internet kursiert, wieder entfernt zu bekommen. Eine weitere Problematik für die Betroffenen ist die juristische Aufarbeitung: Die Erstellung und Verbreitung von Deepfakes oder Deep Nudes wird bislang nur als „Antragsdelikt" verfolgt. Das bedeutet, dass Betroffene selbst die Initiative ergreifen und sich einen Rechtsbeistand suchen müssen; dieser muss dann persönlich vorfinanziert werden und es besteht die Gefahr, dass die Betroffenen auf den Kosten sitzen bleiben. Immerhin existieren mittlerweile Organisationen wie HateAid, die Betroffene beraten und unterstützen (vgl. Hate Aid 2024). Eine technische Initiative stellt „Am I In Porn?" dar. Dabei handelt es sich um eine Suchmaschine, die auf Basis eines hochgeladenen Fotos ermittelt, ob sich auf bekannten Pornografie-Plattformen Werke mit der abgebildeten Person befinden (vgl. Wikipedia 2023). Dieser Ansatz ist aber durchaus nicht unproblematisch, da nicht sichergestellt werden kann ob Nutzende nicht Fotos von Dritten zur Suche hochladen.

Eine präventive Maßnahme stellen für KI nicht mehr nutzbare, sogenannte vergiftete Fotos dar. Wie das funktioniert, wird im nächsten Abschnitt erläutert.

[1] Vgl. den Art. 35 Abs. 1 lit. k Digital Services Act (DSA) (Europäische Union 2022)



Schutz vor KI-Datenabfluss oder -Weiterverwendung

Der KI-Hype geht an traditionellen Datenverarbeitungssystemen nicht spurlos vorbei. Immer mehr Software-Hersteller integrieren KI-Funktionalität in ihre Produkte. Für Anwender*innen sind die Konsequenzen der neuen Funktionen wenig transparent.

In sehr vielen Fällen kommen KI-Erweiterungen zum Einsatz, die nicht vertraulich und geschützt auf dem lokalen Endgerät laufen, sondern die bearbeiteten Dokumente an externe Rechenzentren übermitteln. Erst dort sind dann die KI-Systeme beheimatet, die die Daten zwischenspeichern und verarbeiten.

Aus Sicht von Anwender*innen ist also erhöhte Vorsicht geboten, wenn eingesetzte Software plötzlich neue Komfortfunktionen anbietet. Analysen von IT-Expert*innen haben zudem ergeben, dass Hersteller wie Microsoft die Aktivierung von KI-basierten Erweiterungen entweder sehr kurzfristig ankündigen oder die Kund*innen über die Umsetzung im Unklaren lassen.[2] Eine weitere Problematik besteht darin, dass die zusätzlichen KI-Funktionen von Nutzer*innen ohne Admin-Rechte aktiviert werden können, sodass Arbeitgeber und deren IT-Zuständige aufwändig nacharbeiten oder zu den Herausforderungen schulen müssen. Im November 2023 hat Microsoft zudem angekündigt, den hauseigenen KI-Assistenten „Copilot“ auch für das technisch angegraute Windows 10 via Update-System auszurollen (vgl. heise online 2023b). Im Frühjahr 2024 folgte die sehr umstrittene „Recall“-Funktion, die automatisch und in kurzen Abständen Bildschirmfotos auf den Endgeräten der Nutzenden macht und damit, KI-basiert, Recherchen zu früheren Tätigkeiten ermöglichen soll.

Die Trainingsdaten für KI-Systeme entstammen überwiegend aus frei zugänglichen Quellen im Internet. Neben bekannten, hochwertigen und vielsprachigen Wissensdatenbanken wie der Wikipedia durchforsten sogenannte Crawler-Bots das ganze Internet, um neue Inhalte für die KI-Systeme zusammenzutragen. Daher stellt sich für Organisationen und Privatpersonen die Frage, inwieweit die eigenen Veröffentlichungen Teil dieses Datenschatzes sein sollen.

Es existieren hierzu mehrere Abwehr- und Einschränkungsmechanismen:

- Crawler-Bots von großen KI-Systemen können via „robots.txt“ ausgeschlossen werden (vgl. OpenAI Platform o. J.). Dabei handelt es sich um eine lange etablierte Technik, die auf freiwillige Kooperation setzt und ursprünglich für Suchmaschinen-Bots entwickelt wurde. In einer auf dem Server zu hinterlegenden Datei sind maschinenlesbare Ausschluss- oder Zulassungsdirektiven hinterlegt. Somit kann man beispielsweise festlegen, dass Texte zum Training genutzt werden dürfen, aber Fotos und Bilder für die KI tabu sein sollen. Leider halten sich nicht alle Bot-Hersteller an diese Regeln.
- Bots sind als Teilnehmende im Internet mit einer oder mehreren IP-Adressen verknüpft. Zusammenstellungen dieser IP-Adressen finden sich im Web. Als Website-Betreiber*in kann man diese Informationen nutzen und IP-Adressen vom Seitenabruf aussperren. Allerdings ändern sich die IP-Adressen immer wieder und gerade unbekanntere Bots sind damit nicht erfasst.

[2] Microsoft hat offenbar bereits 2023 in bestehenden, älteren Office-Installationen über das Update-System KI-Funktionen eingebaut und aktiviert (vgl. heise online 2023a).

- Nach dem deutschen Urheberrecht kann man sich das Recht für kommerzielles Text- und Data-Mining im Sinne von § 44b UrhG vorbehalten und stattdessen einen Lizenzerwerb anbieten. Eine Veröffentlichung eines entsprechenden Passus genügt (theoretisch) – ob sich Bots daran halten, darf stark angezweifelt werden.
- Es existieren auch Empfehlungen, in veröffentlichte Internetseiten große Passagen von Lorem-Ipsum-Texten einzufügen, z. B. in unsichtbarer Schriftfarbe im Hintergrund. KI-Crawler sollen die Inhalte dadurch eher als irrelevant verwerfen. Möglicherweise schadet man damit aber auch seiner Suchmaschinen-Wertung.
- Ein neuerer Ansatz, um multimediale Daten für KI-Systeme unbrauchbar zu machen, ist die „Vergiftung“ durch Störddaten. Dabei handelt es sich um für Menschen nicht wahrnehmbare Veränderungen, die die KI-Erkennung täuschen. Im Augenblick existieren hier vor allem im Bereich der Bildbearbeitung verschiedene Forschungsprojekte wie Photoguard oder Glaze/Nightshade, die ihre Software als Open-Source-Anwendung zur Verfügung stellen.

i

github.com/MadryLaphotoguard

glaze.cs.uchicago.edu

nightshade.cs.uchicago.edu

06

Künstliche Intelligenz verstehen und nutzen

Kann man KI-generierte Inhalte erkennen?

Es stellt sich die Frage, ob man KI-Werke von „traditionell“ erstellten Werken unterscheiden kann. Die mitunter erschreckende Erkenntnis ist: Es wird zunehmend schwieriger.

Die Kennzeichnungspflicht zielt dabei auf zwei Methoden ab: Einerseits sollen künstlich generierte Werke für die Betrachter*innen im Zweifelsfall sofort erkennbar sein. Technisch werden den erstellten Dateien überdies kennzeichnende Wasserzeichen und Meta-Daten beigefügt sein, die sich Idealfall nicht mehr entfernen lassen. Im Augenblick bestehen mehrere konkurrierende Systeme wie SynthID von Google oder das CR-Transparenzsystem der „Coalition for Content Provenance and Authenticity“ (C2PA) von Adobe, ARM, Intel Microsoft und Truepic (SynthID: www.deepmind.com/synthid, CR: https://contentcredentials.org). Welche Systeme sich auf lange Sicht durchsetzen werden, ist im Augenblick schwer abzusehen.

Darüber hinaus existieren Prüftools, die sich daran versuchen, künstlich generierte Werke automatisch zu erkennen. OpenAI, die Firma hinter ChatGPT, Dall-E und Sora, hat ihr eigenes Prüftool wieder zurückgezogen, da sie die Zuverlässigkeit der automatisierten Erkennung grundsätzlich nicht sicherstellen können (vgl. OpenAI 2023).

Was also tun, wenn sich die Einsteller*innen nicht an eine Kennzeichnungspflicht halten?

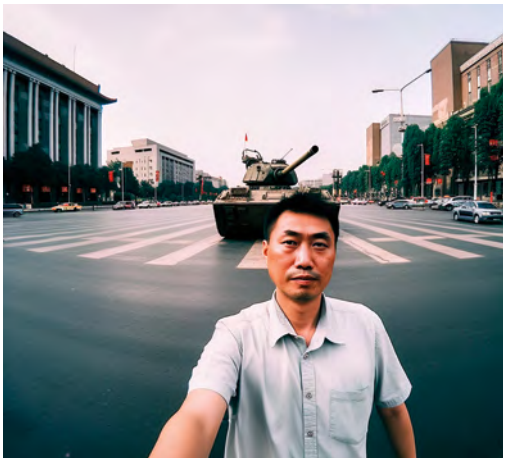
Folgende Tipps können helfen, KI-generierte Inhalte dennoch zu erkennen:

Texte: Unter den KI-generierten Werken sind Texte am schwersten zu erkennen, denn sie werden häufig noch von einem menschlichen Akteur nachbearbeitet oder ergänzt. Rein computergenerierte Texte wirken meist geschliffener und etwas oberflächlicher als menschlich erstellte Schriften. Auch das Sprachniveau passt mitunter nicht zum angeblichen Urheber.

In Chats schnell auffällig sind vor allem die gleichbleibende Freundlichkeit und das Fehlen von Tippfehlern. Bei Logikfragen oder mathematischen Aufgaben versagen KI-Systeme schnell. Neueste gesellschaftliche „Aufreger“ oder Nachrichten sind ihnen zudem meist unbekannt.

Fotos/Bilder: Besonders achten sollte man auf Gliedmaßen wie Finger, Arme oder Beine. Nicht selten tauchen diese in falscher Anzahl oder an falscher Stelle auf. Augen werden gerade bei kleiner dargestellten Personen häufig noch etwas verzerrt generiert. Schriftzüge erscheinen häufig verschwommen oder ergeben bei näherer Betrachtung keine sinnvollen Wörter. Kleinere Aspekte wie Flaggen oder Embleme werden ebenfalls häufig inkonsistent generiert. Neben den Fehlern gibt es auch die Gegenseite: Ebenfalls sind übermäßig perfekt in Szene gesetzte Fotos ein starkes Indiz für eine computer- oder KI-unterstützte Erstellung.

Die Bilder wurde im Stil eines Animationsfilms mit Stable Diffusion generiert. Es sind Beispiele für immer wieder auftretende Fehler bei KI-generierten Bildern.

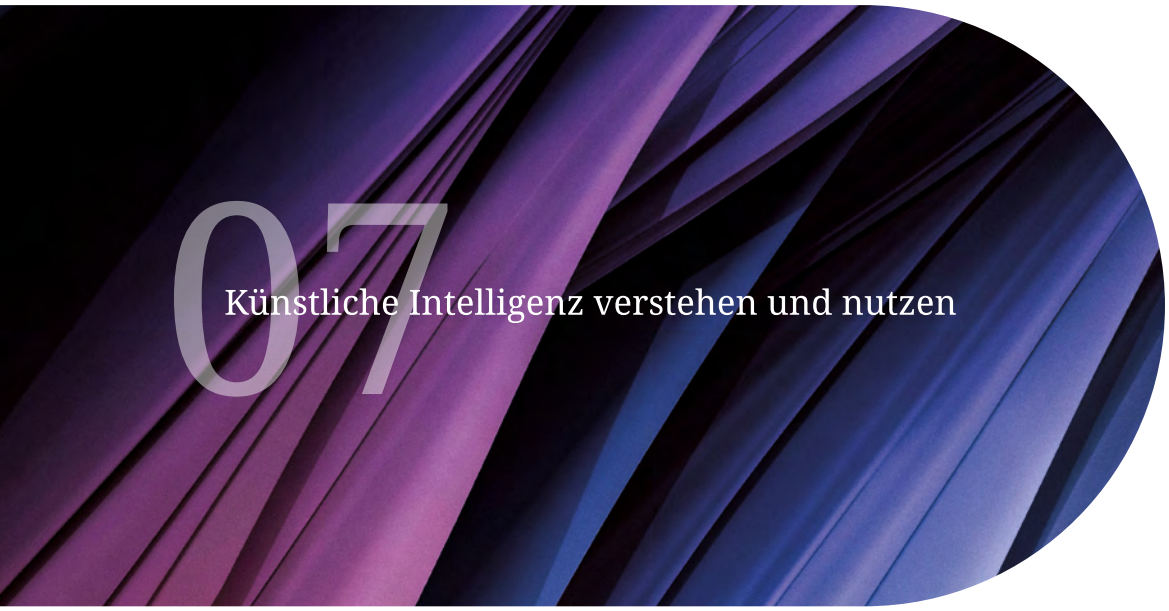


Von den Protesten rund um den Tiananmen-Platz in Beijing im Jahr 1989 gibt es ein weltberühmtes Foto, bei dem sich ein Mann vor einen Panzer stellt und ihn damit aufhält. Wer dieses Foto bei Google sucht(e), erhielt aber in letzter Zeit vermehrt KI-generierte Bilder, die ein vermeintliches „Selfie" von dem Vorfall zeigen. Hier ein Beispiel.

Quelle: https://petapixel.com/2023/09/27/ai-image-of-tiananmen-squares-tank-man-rises-to-the-top-of-google-search

Audio/Sprache: Besonders auffällig sind unnatürliche Betonungen oder die falsche Aussprache von Fremdwörtern. Atemgeräusche oder Rauschen taugen mittlerweile nicht mehr so gut als Indiz, da diese häufig ebenso generiert werden können. Metallischer Klang kann ein Indiz sein, ist aber oft auch der Übertragungskompression geschuldet. Weitere Auffälligkeiten können übermäßig perfekte Satzkonstruktionen und das Fehlen von dialektalen oder Akzent-Charakteristika der natürlichen Person. Bei einer Live-Unterhaltung kann auch zeitliche Verzögerung oder die unnatürliche Reaktion auf Unterbrechungen den entscheidenden Hinweis geben. Inhaltlich gibt es zudem die Möglichkeit nach einem Thema zu fragen, das ein Faker nicht korrekt beantworten kann.

Video: Hier fallen am ehesten Gesichtsmanipulationen und unnatürliche Bewegungen auf. Die Mimik kann eingeschränkt sein, Konturen vor allem im Gesicht und Übergänge wirken leicht verwaschen. Starke Drehbewegungen führen häufig zu Auffälligkeiten, weil KI-Systeme Dreidimensionalität noch nicht richtig beherrschen. In Videosequenzen ändern Akteure oder Gegenstände mitunter ihre Gestalt.



Nachhaltigkeit und Ressourcenverbrauch

KI-Systeme bauen auf neuronalen Netzen auf. Diese sind aufgrund ihrer Struktur um ein Vielfaches energiehungriger als reine Datenbankabfragen, wie sie bei „traditionellen“ Suchmaschinen vorherrschten.

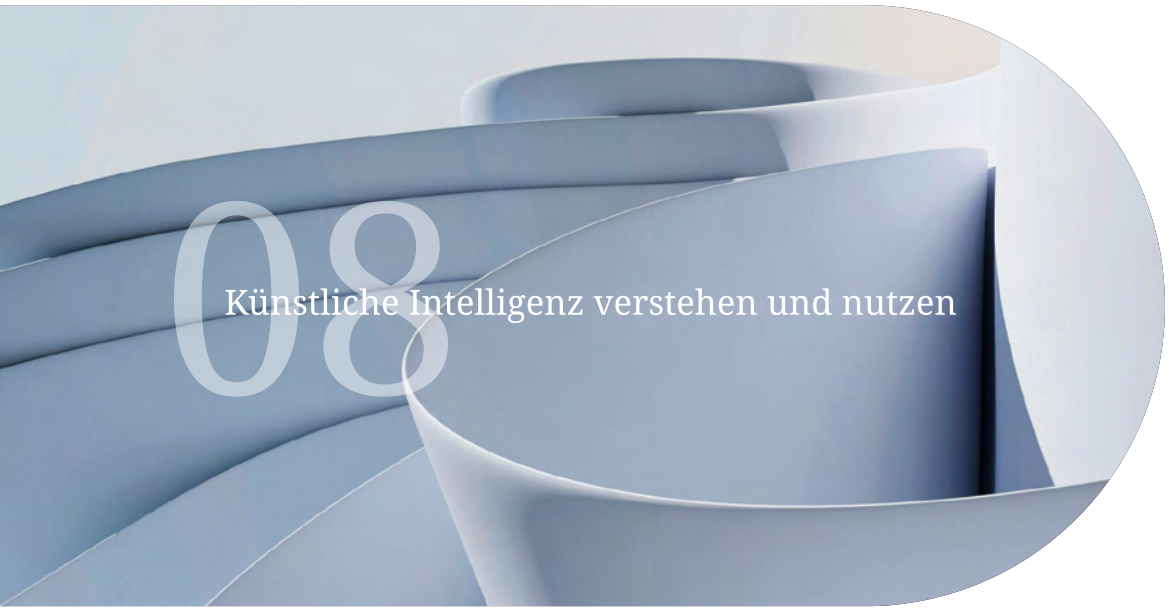
Bereits das Training ist aufwändig: Zum Erlernen von Informationen genügt einem KI-System nicht der eine wissenschaftliche Artikel zum Thema, sondern es muss ein Netz an miteinander verbundenen und sich überlagernden Dateninseln erstellt werden. Am Ende des Trainings können derzeitige KI-Systeme wie Large Language Models nicht mehr zurückführen, wodurch sie einzelne Informationen erlernt haben.

Aufgrund der Entwicklung spezialisierter Rechenchips ist davon auszugehen, dass der Stromverbrauch sich etwas verringern wird, wie das z. B. beim chip-unterstützten Video-Dekodieren bereits in jedem Endgerät der Fall ist. Der Einsatz von KI-Systemen wird dennoch dauerhaft mehr Energiebedarf mit sich bringen als reine Datenbankabfragen. Abhängig von einem Recherchethema kann man als Nutzer*in also bewusst nachhaltig entscheiden.

Die Frage der Nachhaltigkeit von KI-Systemen beschäftigt auch Firmen und Organisationen. Strombedarf ist dabei nur eines der Themen. Soziale und gesellschaftliche Nachhaltigkeit spielt eine größer werdende Rolle. Einzelne große Software-Hersteller im Open-Source-Bereich sind dabei Vorreiter: Die Nextcloud GmbH hat beispielsweise bereits im Sommer 2023 ein Ampelsystem für KI-Erweiterungen und -Funktionen veröffentlicht (vgl. Nextcloud 2023). Unter dem Label „Ethical AI“ (Ethische KI) zeigt eine einfach verständliche Ampel, wie es um Privatsphäre und Vertraulichkeit der verarbeiteten Daten, um Diskriminierung und Bias sowie um den Energieverbrauch der jeweiligen Lösungen bestellt ist. System-Admins und Nutzer*innen können so relativ schnell entscheiden, welche der Optionen sie einsetzen möchten. Oft stehen mehrere KI-Systeme parallel zur Auswahl.

„Auf eine Selbstregulierung in Unternehmen kann man derzeit allerdings nicht uneingeschränkt vertrauen.“

Auf eine Selbstregulierung in Unternehmen kann man derzeit allerdings nicht uneingeschränkt vertrauen. Einer der größten KI-Anbieter, Microsoft, hat beispielsweise im März 2023 sein „Ethics and Society Team“ abgeschafft (vgl. Ziegener 2023). Auch Google verhält sich eher zögerlich: Die ab November 2023 geltenden Transparenzregeln schreiben lediglich vor, dass bezahlte Wahlwerbung mit KI-Elementen gekennzeichnet werden müsse, andere Inhalte hingegen nicht (vgl. Google 2023).



08 Künstliche Intelligenz verstehen und nutzen

Für welche Zwecke kann ich KI-Anwendungen und -Dienste sinnvoll nutzen?

Sofern der Einsatz von KI-Systemen über ein Experimentieren und Kennenlernen hinausgeht, sollte die Frage nach Bedarfen und gewünschten Einsatzszenarien gestellt werden.

Derzeit können KI-Systeme sinnvoll eingesetzt werden in folgenden Bereichen:

Bildbearbeitung

- Bilder und „Fotos“ können per KI generiert, vorhandene erweitert („Outpainting“) und retuschiert („Inpainting“) werden.
- Vorhandene Bilder können in einen anderen Stil oder ein anderes Setting überführt werden. Aus einfachen Skizzen können detaillierte und fotorealistische Darstellungen generiert werden.

- KI kann darüber hinaus versuchen, Bilder in schlechter Qualität auf eine bessere Qualität hochzurechnen („Upscaling“).
- Bekannte Tools sind hier z. B. Midjourney, DALL-E, Stable Diffusion und Adobe Firefly.

Texterstellung und -bearbeitung

- Erstellung eines Textes inkl. inhaltlicher Recherche auf Basis einer einfachen oder ausführlichen Aufgabenstellung (Prompt)
- Ausformulieren von Textskizzen und Erweitern von vorhandenen Textpassagen
- Zusammenfassen von Text
- Übersetzen in andere Sprachen oder in andere Sprachniveaus, Soziolekte und Dialekte
- Bekannte Tools sind z. B. OpenAIs ChatGPT, Claude Anthropic, Google Gemini, Meta Llama, Mistral und Microsoft Copilot. Auf Übersetzungen spezialisierte Tools, die auch Texte in Fotos erkennen, sind Deepl.com und Google Translate.

Audio-Generierung und -verarbeitung

- Erstellung von Geräuschen und Musik auf Basis einer textlichen Aufgabenstellung (Prompt)
- Diktate und Transkription von Audio (Speech-To-Text)
- Vorlesen vorhandener Texte (Text-To-Speech) und das Nachahmen von menschlichen Stimmen
- Bekannte Tools sind Stable Audio und für die Stimmenimitierung Fliki, ElevenLabs und Murf.ai; zur Texttranskription bietet sich u. a. Whisper an.

Video-Generierung und -bearbeitung

- Erstellung von kurzen Videosequenzen auf Basis einer textlichen Aufgabenbeschreibung
- Austausch von Gesichtern in Videos
- Bekannte Tools sind RunwayML, Stable Video Diffusion und Sora sowie Filter in Social-Media-Apps.

Muster- und Objekterkennungen

- Gesichts- und Objekterkennung in Fotos und Videos sowie für Authentifizierungsverfahren („Windows Hello“)
- Emotions- und Aufmerksamkeitserkennung (bei Smartphones oder in Kfz-Assistenzsystemen)
- „Smarte“ Mail-Postfächer
- Hintergrund-Anpassung in Video-Calls
- Erkennung von verdächtigen Login- oder IT-Angriffsversuchen

- Analysetechnik in der Medizin (z. B. Erkennung von Auffälligkeiten in CT-/MRT-Aufnahmen)

Tutoring- und Assistenzsysteme

- KI-Tools, die speziell darauf optimiert sind Lernerfolge zu fördern, indem sie individuell angepasste Aufgaben für die Lernenden zusammenstellen und Feedback geben
- Kurzfristige oder auf längere Zeitspannen angelegte Assistenzsysteme und Apps
- Tools gibt es bei spezialisierten Anbietern oder können durch die „GPTs“ bei ChatGPT selbst erstellt werden.

Kann man KI lokal und ohne Datenabfluss nutzen?

Die meisten Systeme Künstlicher Intelligenz werden online bei den jeweiligen Herstellern genutzt. Die Anbieter kümmern sich dabei um die Bereitstellung der Infrastruktur und die Pflege des Systems. Gleichzeitig bedeutet dies aber auch, dass alle Nutzereingaben und KI-Ergebnisse auch dem Anbieter bekannt sind.

Allerdings existiert auch eine große Auswahl an frei verfügbaren und rein lokal nutzbaren KI-Systemen. Während die Installation und Konfiguration zunehmend einfacher wird, ist man nach wie vor auf starke Hardware angewiesen. Insbesondere die Rechenleistung von Grafikkarten werden in KI-Systemen genutzt, um die Wartezeit bis zur Ergebnisausgabe zu verkürzen.

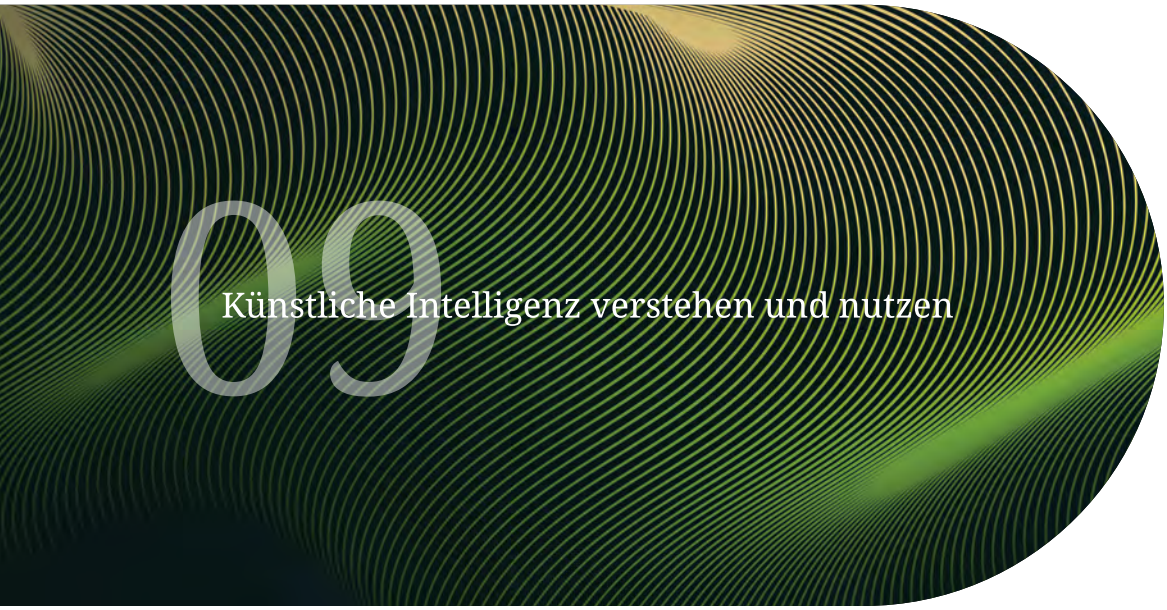
Beispiele für Open-Source-KI:

- Mistral, Meta Llama, Alpaca, Google Gemma (Text/Chat/Übersetzung)
- Stable Diffusion (Bild- und Video)
- OpenAI Whisper (Audio-Transkription)
- Firefox Translate/Nextcloud Translate (Textübersetzung)

Einige dieser Tools können auf einem Gerät installiert und dann per Fernzugriff gemeinsam in einem lokalen Netzwerk genutzt werden, also z. B. innerhalb einer Bürogemeinschaft oder einer Bildungseinrichtung. Dies reduziert den Bereitstellungsaufwand.

Die freie KI-Verwaltung „Pinokio“ bietet eine kuratierte Übersicht von lokal und datenschutzkonform betreibbaren KI-Tools und den dazugehörigen Ein-Klick-Installationen, ist aber an einigen Stellen eher für einen experimentellen Einsatz konzipiert.

pinokio.computer



Checkliste für den KI-Einsatz

Die folgende, in Sektionen untergliederte Checkliste dient der strategischen Planung und Nutzungsumsetzung für KI-Systeme sowohl im Büro- und Verwaltungsalltag als auch in der Bildungsarbeit.

Während sich die Technik stetig und schnell weiterentwickelt, soll die Checkliste einen möglichst beständigen Rahmen geben, um die Nutzung nachhaltig zu ermöglichen. Sie entstand durch Anregungen von Datenschutzbeauftragten und aus dem Austausch mit Fachkräften in der Bildungsarbeit.

1. Vorbereitung: Ziele definieren und Zuständigkeiten klären

- Der KI-Einsatz sollte sich an Bedarfen orientieren. Dabei ist zu bedenken, dass KI-Anwendungen systemische Probleme einer Organisation nicht lösen können. Der Einsatz kann also lediglich als Unterstützung, Tätigkeitserweiterung oder in manchen Bereichen als teilweiser Ersatz für vorhandene Abläufe gelten. Eine menschliche Kontrolle bleibt dennoch dauerhaft notwendig.
- Um dem Team einen sicheren Einsatz von KI-Tools zu ermöglichen, sollte ein gemeinsames Regelwerk aufgestellt und gut zugänglich dokumentiert werden. Ohne vereinbarte Regeln ist davon auszugehen, dass sich Team-Mitglieder auf eigene Faust der neuen Werkzeuge bedienen. Unter Umständen steht ein Arbeitgeber dennoch für abgeflossene Daten oder entstandenen Schaden in der Verantwortung und muss Haftung übernehmen.
- Bei der Erstellung eines Regelwerks sollten Antidiskriminierungsbeauftragte, Betriebs- und Personalräte sowie Datenschutzbeauftragte mit einbezogen werden. Je nach Organisationsgröße kann hierfür sogar die Pflicht der Einbeziehung bestehen. In sehr seltenen Fällen kann zudem eine Datenschutz-Folgenabschätzung notwendig sein.
- Das Regelwerk kann als gemeinsam entwickelte Richtlinie, als Teamvereinbarung, Betriebsvereinbarung oder als Dienstanweisung veröffentlicht werden.
- Im Regelwerk sollte aufgeführt werden, für welche Zwecke KI-Anwendungen genutzt werden dürfen und welche KI-Anbieter konkret dafür zulässig sind. Diese Liste kann als Anhang zu den Regeln ggf. häufiger ergänzt und überarbeitet werden.
- Damit Team-Mitglieder die Regeln zuverlässig befolgen, sollte klar gemacht werden wofür die Regeln gut sind. Hinweise auf Schutz von Vertraulichkeit, sichere Erfahrungs- und Experimentierräume und der Schutz vor langfristigen Folgen sind sinnvolle Ausgangspunkte.
- Neben der Positivdarstellung sind Beispiele für Grenzen hilfreich, die nicht überschritten werden dürfen.
- Das Regelwerk sollte Verfahren beschreiben, wie Probleme und neue Anwendungsszenarien in Zukunft berücksichtigt werden.

2. KI-Positiv- und Negativliste führen und Team schulen

- Die Aufstellung von verbindlichen Regeln ist ein guter Startpunkt. Damit Team-Mitglieder und Teilnehmende die Vereinbarungen auch einhalten, ist eine Sensibilisierung notwendig. Schulungen helfen dabei, das Bewusstsein für die Regeln und deren Begründung wachzuhalten.
- Die Regeln können beispielhaft in einer Positiv- und Negativliste dargestellt oder konkretisiert werden. Auch können Beispielformulierungen zu Tools helfen, die Nutzungsabsprachen intuitiv verständlich zu machen.

  Beispiele:

  Mit ChatGPT dürfen Sie Pressemitteilungen überarbeiten.
  Mit ChatGPT dürfen Sie nicht interne Sitzungsprotokolle überarbeiten.
  Mit Midjourney dürfen Sie Illustrationen für Online-Beiträge erstellen.
  Mit Midjourney dürfen Sie nicht echte Menschen abbilden.
  KI-Bilder dürfen keine Menschen in akuten Notsituationen darstellen.
- Sofern technische Kompetenz vorhanden, kann eine KI-Anwendung lokal vorgehalten werden. Daraus können sich weitere Datenschutzfragen ergeben, die bei externen Dienstleistern möglicherweise nicht auftreten (z. B. mehr/versehentlicher Einblick durch Kolleg*innen).
- Für experimentelle Nutzung können darüber hinaus Tools aufgelistet werden, die für die alltägliche Nutzung nicht alle Qualitäts- und Vertraulichkeitskriterien erfüllen, wie z. B. anonym nutzbare Tools von https://deepai.org oder die Chatbot-Arena unter https://lmarena.ai.
- Die Aufgabenstellung an KI ist trotz der natürlichsprachlich zu formulierenden Prompts gerade am Anfang mit Frustrationserlebnissen verbunden. Häufig benötigt man mehrere Anläufe und Erfahrung, wie man die KI dazu bringt, erhoffte Ergebnisse zu produzieren. Daher bietet es sich an, einen Ort zu definieren, an dem erfolgreiche Prompts zur Anregung und Anpassung in der Organisation hinterlegt werden.
- Für weitergehende Fragen sollte in der Organisation eine konkrete Ansprechperson und eine wiederkehrende Gesprächsrunde festgelegt werden, damit Bedarfe und Unsicherheiten nicht zu unvorhergesehenen Aktivitäten führen.

deepai.org

lmarena.ai

3. Accounts einrichten und konfigurieren

- Für die meisten kommerziellen KI-Systeme müssen sich Nutzende registrieren. Hier gilt: Auch wenn es für die ersten Gehversuche verlockend erscheint, so sollten berufliche und private Accounts für KI-Anwendungen von Anfang an strikt voneinander getrennt werden. Nur so ist gewährleistet, dass keine Vermischung von Inhalten in den KI-Ergebnissen stattfindet.
- Beschäftigte sollten nicht eigenständig und unter Verwendung privater Daten ein Konto erstellen. Denn so würde beim KI-Dienst im Laufe der Nutzungszeit ein umfassendes Interessen- und Tätigkeitsprofil zu den jeweiligen Personen entstehen. Nach Möglichkeit sollten diese Accounts ebenso nicht die Namen einzelner Personen enthalten; es bieten sich daher besser eine eigens dafür eingerichtete, funktionale E-Mail-Adresse an, anstatt bestehender, häufig namensbasierter Adressen.
- Teilweise werden zur Registrierung Mobilfunknummern vom KI-Hersteller verlangt. Soweit diese Angabe nicht vermeidbar ist, empfiehlt es sich hier, ein dienstliches Telefon zu benutzen. Arbeitgeber müssen diese Möglichkeit zur Verfügung stellen, wenn sie die Nutzung des jeweiligen Dienstes wünschen.
- Wie für alle digitalen Dienste gilt: Accounts sollten mit starken Passwörtern abgesichert werden, sobald darüber personenbezogene oder vertrauliche Daten verarbeitet oder eingesehen werden können. Verfahren mit Zwei-Faktor-Authentifizierung können dabei helfen, unberechtigte Zugriffsversuche rechtzeitig zu erkennen.
- In den Nutzungsbedingungen behalten sich Hersteller von KI-Systemen häufig vor, alle getätigten Eingaben zum weiteren Training auszuwerten. Dadurch erhalten nicht nur Mitarbeitende des KI-Systems Kenntnis der verarbeiteten Inhalte, auch können Dritte nach dem Training Inhalte dann absichtlich oder zufällig im KI-System erfragen. Aus diesem Grund sollte man die Weiterverarbeitung der eingegebenen Daten ablehnen. Dies geschieht über ein ausdrückliches Opt-Out: Insbesondere die Verwendung der Daten zu Trainingszwecken sollte abgelehnt werden. Je nach Dienst muss hierfür möglicherweise ein spezielles Vertragsmodell gebucht werden.[3]

[3] Bei OpenAI/ChatGPT ist das Opt-out derzeit möglich über die Einstellungen unter EINSTELLUNGEN → DATENKONTROLLEN → MODELLVERBESSERUNG.

4. Nutzung: Nichts Vertrauliches eingeben oder hochladen

Gerade Accounts für KI-Chatbots bieten erhebliches Missbrauchspotenzial. Sie beinhalten meist eine Historie aller Aktivitäten und Chatverläufe einer konkret identifizierbaren Person. Mehr noch als bei KI zur Bild- oder Mediengenerierung finden sich daher in diesen Verläufen häufig auch Interna.

In den falschen Händen kann dieses Wissen anderweitig genutzt oder mit zusätzlichen Daten angereichert werden, um diese für Phishing- oder Social-Engineering-Angriffe zu missbrauchen. Der Diebstahl einer digitalen Identität kann schwerwiegende Folgen haben (z. B. leergeräumte Bankkonten oder Datenverlust).

- **Historie (zeitweise) deaktivieren:** KI-Dienste bieten in der Regel an, bisherige Eingaben zu speichern, um den Dialog zu einem Thema an einem späteren Zeitpunkt wieder aufnehmen zu können. Diese Historie kann bei vielen Diensten vorübergehend oder dauerhaft deaktiviert werden. Bei der gemeinsamen Nutzung eines Accounts durch mehrere Personen ist es sogar ratsam, die Historie generell abzuschalten und nur in bewusst ausgewählten Fällen zu aktivieren: Versehentliche Datenabflüsse oder kommunikative Missverständnisse können dadurch reduziert werden.
- **Keine Datenverarbeitung ohne Rechtsgrundlage:** Seit Einführung der DSGVO ist den meisten Akteur*innen bewusst, dass man für die Verarbeitung personenbezogener Daten die Einwilligung von betroffenen Personen benötigt. Dies gilt auch für die Datenverarbeitung mittels KI-Systemen. Daten von Beschäftigten, Seminar-Teilnehmenden, Kund*innen, Geschäftspartner*innen oder Dritten sind daher tabu. Es wird sich kaum eine belastbare Rechtsgrundlage finden lassen, die die Eingabe dieser Daten in ein online-basiertes KI-System rechtfertigt.
- **Personenbeziehbarkeit der Eingaben vermeiden:** Eingaben, die unter Umständen auf konkrete Personen bezogen werden können, sollten ebenfalls vermieden werden. Die Anzahl von Akteur*innen in einem Tätigkeitsfeld ist oft überschaubarer als man vermutet. Es genügt daher nicht, konkrete Daten wie Namen und Anschriften aus der Eingabe zu entfernen. Auch aus dem Zusammenhang lassen sich möglicherweise Rückschlüsse auf Autor*innen und Betroffene ziehen. Die Stärke von KI-Anwendungen besteht darin, Querbezüge auch aus unstrukturierten Daten herzustellen; daher ist die Gefahr einer ungewollten Informationsweitergabe hier besonders hoch.

- **Opt-out des Trainings:** Die Anbieter von KI-Diensten räumen sich in den AGB fast immer eine Weiternutzung der eingegebenen Daten ein: Diese werden zur Qualitätsanalyse wie auch zum weiteren Training verwendet. In letzterem Fall würden also eingegebene oder hochgeladene Informationen auch bei anderen Nutzer*innen als Teil einer Ausgabe erscheinen können. Selbst Bezahl-Accounts nehmen die Anbieter davon nicht unbedingt aus. Daher gilt: Auf keinen Fall sollten in online-basierten KI-Diensten personenbezogene oder personenbeziehbare Daten eingegeben werden.
- **Fragestellungen nicht auf personenbeziehbare Antworten formulieren:** Auch wenn das Prompt (Eingabebefehl) selbst keine vertraulichen Daten verwendet, kann die Aufgabenstellung selbst eine KI-Antwort mit Personenbezug auslösen. Ausgaben können eine diskriminierende Wirkung entfalten, wenn diese Ergebnisse an anderer Stelle, z. B. als Entscheidungsgrundlage, weiterverwendet werden. Fragen nach Zielgruppen oder idealtypischen Personen gilt es daher zu vermeiden.
  Weiterhin können solche Abfragen dazu führen, dass die KI unter Umständen frühere Konversationen oder Informationen aus dem Internet einbezieht und nicht erwünschte oder unrichtige Querbezüge herstellt.[4]

5. Umgang mit Ausgaben der KI

Heutige KI-Systeme sind darauf getrimmt, überzeugende Ergebnisse zu produzieren. Dabei kommt es dann nicht auf eine inhaltliche Richtigkeit an – diese existiert so direkt für ein KI-System gar nicht –, sondern auf die höchsterrechnete Wahrscheinlichkeit einer Ausgabeoption. Wenn diese Ausgabe nicht den Fakten entspricht, nennt man das eine „KI-Halluzination“.

Ergebnisse sollten daher stets geprüft werden auf

- **Richtigkeit:** Systeme Künstlicher Intelligenz greifen häufig auf ältere Informationsstände zurück, da die Hersteller einen gewissen Zeitraum zum Trainieren und Nachjustieren benötigen. Die Einbeziehung von tagesaktuellen Quellen führt häufiger zu Fehlinterpretationen durch das KI-System.

[4] KI-Querbezüge können rufschädigende Auswirkungen haben. Microsofts Copilot konnte bspw. Mitte 2024 nicht zwischen einem Gerichtsreporter und den Straftätern aus den berichteten Fällen unterscheiden (vgl. Beschoner 2024).

Aus diesem Grund sind aktuelle Informationen oft nicht verfügbar oder schwerer abzurufen als in regulären Suchmaschinen. Zudem sind die Quellen, die von KI-Systemen verwendet werden, in der Regel nicht transparent, was die Überprüfbarkeit und Verlässlichkeit der bereitgestellten Daten erschwert.

- **Bias (inhaltliche Verzerrung):** Verzerrungen können auftreten, weil Trainingsdaten nur aus bestimmten Quellen stammen, nur bestimmte Sprachniveaus berücksichtigen oder weil überwiegend einseitige Informationen vorhanden waren. Ein Beispiel hierfür ist das sogenannte Survivor-Bias: Weil Berichte überwiegend über erfolgreiche Projekte veröffentlicht werden, erscheint möglicherweise der einseitige Eindruck, dass Projekte stets Erfolgsgeschichten sind. Erkenntnisse zu Faktoren, die ein Scheitern begünstigen, kommen daher vielleicht zu wenig vor. Ebenso zeigen KI-Bots ihre antrainierte Voreingenommenheit darin, dass sie als männlich identifizierten Nutzernamen andere Ergebnisse ausspielen als weiblich identifizierten (vgl. Bastian 2024).
- **Diskriminierung:** KI-generierte Empfehlungen sind aufgrund ihrer Trainingsdaten nicht automatisch neutral und objektiv. Eine diskriminierende Wirkung kann von Empfehlungen ausgehen, selbst wenn sie auf den ersten Blick nicht als „falsch" zu erkennen sind. Die Weiterverarbeitung solcher Ergebnisse wäre unzulässig, wenn sie einen Verstoß gegen das Allgemeine Gleichbehandlungsgesetz (AGG) darstellen. Ebenso könnte die Güterabwägung des Art. 6 Abs. 1 lit. f DSGVO entgegenstehen, weil das KI-System unzulässigerweise sensible Merkmale wie Gesundheits- oder andere Daten ausgewertet haben könnte.

Entscheidungsfindung mit KI?

Möglicherweise möchte man KI-Unterstützung dazu nutzen, um schneller Entscheidungen zu treffen. Diese Idee ist nicht nur aus den genannten Risiken wie KI-Halluzinationen und inhaltlicher Verzerrungen problematisch, auch der Gesetzgeber hat konkrete Grenzen gesetzt: Wenn eine Entscheidung gegenüber einer Person eine rechtliche Wirkung entfaltet – z. B. in einem Bewerbungsverfahren – hat die betroffene Person das Recht, keiner automatisierten Entscheidung unterworfen zu werden.[5]

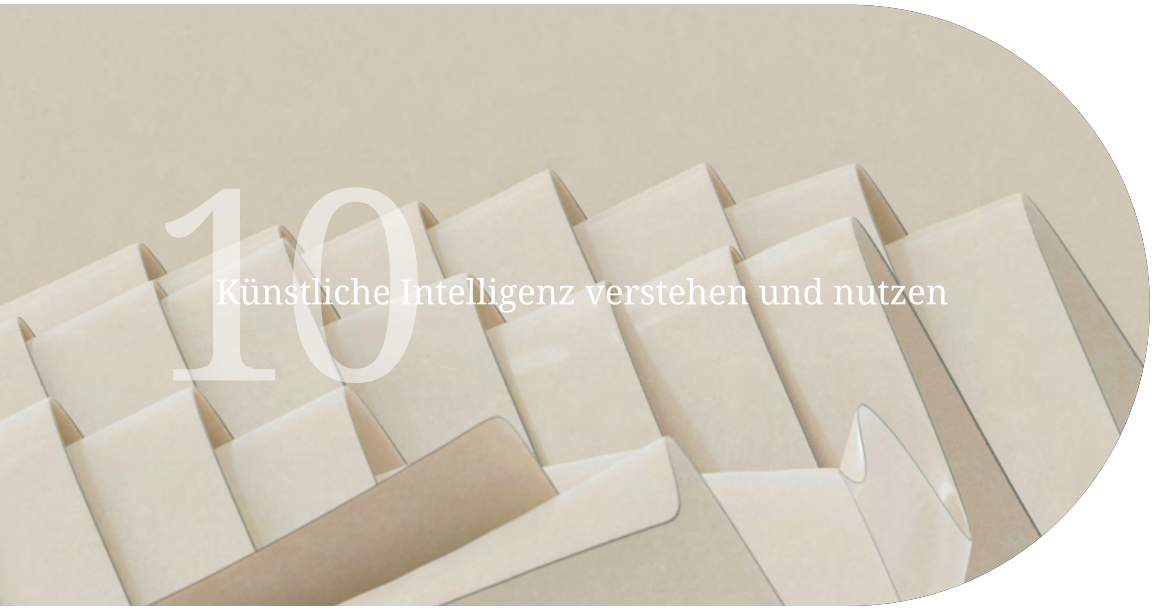
- Entscheidungen sollten grundsätzlich von Menschen getroffen werden.
- Herangezogene KI-Ergebnisse müssen einen Entscheidungsspielraum lassen.
- Ressourcen- und Zeitdruck dürfen nicht dazu führen, dass Ergebnisse ungeprüft übernommen werden.
- KI-gestützte Entscheidungen benötigen einen Faktencheck und die notwendige Dokumentation dieser Prüfung/Argumente.

[5] Vgl. Art 22 Abs. 1 DSGVO.

6. Weiterdenken und Entwicklungen beobachten

- Neben bereits erwähnten Datenschutzaspekten spielen bei der KI-Nutzung auch andere Vertraulichkeitserwägungen wie der Schutz von Metadaten (Gewohnheiten, Orte, Zeiten, Themen und Häufigkeiten) sowie Geschäftsgeheimnisse eine Rolle. In manchen Kontexten müssen auch Schweigepflichten oder das Weitergabeverbot nach dem Sicherheitsüberprüfungsgesetz (SÜG) und ähnliche Regelungen berücksichtigt werden.
- Während KI-generierte Ergebnisse grundsätzlich gemeinfrei sind, können Urheberrecht und Persönlichkeitsrechte bei abgebildeten Figuren oder Personen greifen, selbst wenn sie den KI-Nutzenden nicht bekannt sind. Bislang noch selten umgesetzt besteht aber die Möglichkeit, die Berechtigung zur Verwendung KI-generierter Inhalte bei den Rechteinhaber*innen einzuholen, z. B. über eine Lizenzvereinbarung.
- Ethisch-moralische Fragen der KI-Nutzung gehen über Themen des Datenschutzes und der Vertraulichkeit hinaus. Fiktive Interviews oder die Abbildung von Menschen in Notsituationen können diese ethischen Grenzen überschreiten, wenngleich sie nicht gegen Gesetze verstoßen.[6]
- Die Entwicklung der Technik und des gültigen Rechtsrahmens schreitet schnell voran. Der risikobasierte Ansatz der europäischen KI-Verordnung wird voraussichtlich durch nationale Gesetze ergänzt. Eigene Bedarfe verändern sich. Daher sollten interne KI-Nutzungsabsprachen immer wieder auf den Prüfstand und aktuellen Gegebenheiten angepasst werden.

[6] KI-generierte Fake-Interviews haben bereits zu Gerichtsverfahren und Kündigungen geführt: https://web.archive.org/web/20240717134434/https://www.zdf.de/nachrichten/panorama/schumacher-interview-ki-aktuelle-chefin-100.html



10 Künstliche Intelligenz verstehen und nutzen

Anhang

Literaturverzeichnis

Bastian, Matthias (2024): Männliche Chatbot-Rollen schneiden laut Studie besser ab als weibliche; https://the-decoder.de/maennliche-chatbot-rollen-schneiden-laut-studie-besser-ab-als-weibliche (Zugriff: 14.11.2024)

Beschorner, Markus (2024): KI-Chat macht Tübinger Journalisten zum Kinderschänder. Martin Bernklau wird Opfer der Künstlichen Intelligenz. In: SWR aktuell; www.swr.de/swraktuell/baden-wuerttemberg/tuebingen/ki-macht-tuebinger-journalist-zum-kinderschaender-100.html (Zugriff: 13.11.2024)

Bundesministerium der Justiz/Bundesamt für Justiz (2024a; letzte Änderung): Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz) § 2 Geschützte Werke; www.gesetze-im-internet.de/urhg/__2.html (Zugriff: 13.11.2024)

Bundesministerium der Justiz/Bundesamt für Justiz (2024b; letzte Änderung): Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz) § 60d Text und Data Mining für Zwecke der wissenschaftlichen Forschung; www.gesetze-im-internet.de/urhg/__60d.html (Zugriff: 13.11.2024)

Bundesregierung (2024): Einheitliche Regeln für Künstliche Intelligenz in der EU vom 22.05.2024; https://www.bundesregierung.de/breg-de/themen/digitalisierung/kuenstliche-intelligenz/ai-act-2285944 (Zugriff: 13.11.2024)

California Legislative Information (2024): SB-1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act (2023–2024); https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1047 (Zugriff: 13.11.2024)

Deutscher Bundestag/Wissenschaftliche Dienste (Hrsg.) (2024): Kurzinformation Regulierung von Deepfakes; www.bundestag.de/resource/blob/997214/2bc9f3aeda343398c69deb3c32badf3a/WD-7-015-24-pdf.pdf (Zugriff: 13.11.2024)

Ernst, Sebastian M. (2024): »KI-Hype«, oder: Was sind eigentlich »Transformer«? Vortrag vom 16. März 2024; https://media.ccc.de/v/clt24-266-ki-hype-oder-was-sind-eigentlich-transformer (Zugriff: 13.11.2024)

EU Artificial Intelligence Act (2024): Chapter V: General-Purpose AI Models; https://artificialintelligenceact.eu/de/chapter/5 (Zugriff: 13.11.2024)

Europäische Kommission (2018): Künstliche Intelligenz für Europa. Mitteilung der Kommission vom 25.04.2018; https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=COM%3A2018%3A237%3AFIN (Zugriff: 13.11.2024)

Europäische Kommission (2020): Weißbuch: Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen COM(2020) 65 final. Brüssel; https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0065 (Zugriff: 13.11.2024)

Europäische Kommission (2021): Koordinierter Plan für künstliche Intelligenz; https://digital-strategy.ec.europa.eu/de/policies/plan-ai (Zugriff: 13.11.2024)

Europäische Union (2022): Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste); https://eur-lex.europa.eu/eli/reg/2022/2065/oj/deu (Zugriff: 13.11.2024)

Google (2023): Updates to Political content policy (September 2023); https://support.google.com/adspolicy/answer/13755910?hl=en&ref_topic=29265&sjid=9558797907376646058-NA (Zugriff: 13.11.2024)

Hate Aid (2024): Realität oder Fake? Bedrohung durch Deepfakes; https://hateaid.org/deepfakes (Zugriff: 13.11.2024)

heise online (2023a): Microsoft: AI-Unterstützung in Microsoft Office teilweise schon aktiv; www.heise.de/news/Microsoft-Office-AI-Unterstuetzung-teilweise-schon-aktiviert-8927848.html (Zugriff: 13.11.2024)

heise online (2023b): Microsoft will KI für alle, bringt Copilot auch für Windows 10; www.heise.de/news/Microsoft-will-KI-fuer-alle-bringt-Copilot-auch-fuer-Windows-10-9531350.html (Zugriff: 13.11.2024)

Hutson, Matthew (2023): Wie Staaten weltweit KI in Schach halten wollen; www.spektrum.de/news/wie-nationen-vorhaben-kuenstliche-intelligenz-in-schach-zu-halten/2171376 (Zugriff: 13.11.2024)

Nextcloud (2023): AI in Nextcloud: what, why and how; https://nextcloud.com/blog/ai-in-nextcloud-what-why-and-how (Zugriff: 13.11.2024)

OpenAI (2023): New AI classifier for indicating AI-written text; https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text (Zugriff: 13.11.2024)

OpenAI Platform (o. J.): GPT Actions. Customize ChatGPT with GPT Actions and API integrations; https://platform.openai.com/docs/plugins/bot (Zugriff: 13.11.2024)

The White House (2023): FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence vom 30.10.2023; www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence (Zugriff: 13.11.2024)

Westarp, Rabea (2022): Gefälschte Sexvideos. Es kann jeden treffen; www.tagesschau.de/investigativ/deepfakes-103.html (Zugriff: 13.11.2024)

Wikipedia (2023): Am I In Porn?; https://de.wikipedia.org/wiki/Am_I_In_Porn%3F (Zugriff: 13.11.2024)

Wikipedia (2024): Transformer (Maschinelles Lernen); https://de.wikipedia.org/wiki/Transformer_(Maschinelles_Lernen) (Zugriff: 13.11.2024)

Ziegener, Daniel (2023): Microsoft baut KI-Ethik-Team ab; https://www.golem.de/news/openai-microsoft-baut-ki-ethik-team-ab-2303-172607.html (Zugriff: 13.11.2024

Zum Weiterlesen

- Der AI Act Explorer: https://artificialintelligenceact.eu/de/ai-act-explorer (Zugriff: 14.11.2024)
- Datenschutz-Checkliste zum Einsatz von Chatbots auf Basis von Large Language Models (Hamburgischer Beauftragter für Datenschutz und Informationsfreiheit, 11/2023); https://datenschutz-hamburg.de/news/checkliste-zum-einsatz-llm-basierter-chatbots (Zugriff: 14.11.2024)
- Diskussionspapier: Rechtsgrundlagen im Datenschutz beim KI-Einsatz (Landesbeauftragter für Datenschutz und Informationsfreiheit Baden-Württemberg, 11/2023); www.baden-wuerttemberg.datenschutz.de/rechtsgrundlagen-datenschutz-ki (Zugriff: 14.11.2024)
- Datenschutz-Grundverordnung (DSGVO) mit Erwägungsgründen; https://dsgvo-gesetz.de (Zugriff: 14.11.2024)

Impressum

Herausgeber:
Arbeitskreis deutscher Bildungsstätten e.V. (AdB)
Mühlendamm 3, 10178 Berlin
(030) 400 401 00
info@adb.de
adb.de

Texte: Tim Schrock
Verantwortlich: Ina Bielenberg (V.i.S.d.P.)

Layout, Satz: Willius Design, Berlin
Druck: Pinguin Druck GmbH, Berlin

Bildnachweis:

Christina Wocintechchat on unsplash.com: Titel
Pawel Czerwinski on unsplash.com: Seite 4, 9, 19, 32
Rene Bohmer on unsplash.com: Seite 6
Freepik.com: Seite 11
Maxim Berg on unsplash.com: Seite 13
Steve Johnson on unsplash.com: Seite 16
Guns on unsplash.com: Seite 21
Visax on unsplash.com: Seite 24

Systeme Künstlicher Intelligenz und ihre Rahmenbedingungen unterliegen derzeit einem starken Wandel. Daher aktualisieren wir die Texte regelmäßig in der Online-Version unter

https://www.adb.de/service/publikationen/weitere-veroeffentlichungen/arbeitshilfe-einsatz-von-ki