

# Zukunft gestalten! Mit generativer KI.

Gesellschaftliche Auswirkungen und  
Handlungsansätze



# Vorwort

Generative KI verändert unsere Gesellschaft. Anwendungen wie ChatGPT, Gemini oder DeepSeek erzeugen Texte, erstellen realistisch wirkende Bilder, komponieren Musik, führen menschenähnliche Gespräche und übersetzen nahezu in Echtzeit. Damit erweitert generative KI unsere Möglichkeiten im Arbeits- wie im Privatleben in Bereichen wie Bildung, Medizin oder Unterhaltung. Unternehmen nutzen bereits KI und vermehrt auch generative KI, um Prozesse zu optimieren, Produkte schneller zu entwickeln und kreative Aufgaben effizienter zu gestalten. Kurzum: Generative KI ermöglicht Innovationen in vielerlei Hinsicht!

Ihr Einsatz wirft aber zugleich gesellschaftliche und ethische Fragen auf: Welche Aufgaben wollen wir delegieren, welche Kompetenzen soll der Mensch behalten? Welche Erwartungen stellen wir an die künftige Entwicklung? Was kann generative KI zu Nachhaltigkeit, Guter Arbeit, Gerechtigkeit und einer lebenswerten Zukunft beitragen – und welche Risiken gilt es zu adressieren?

Ein zentrales Thema ist dabei die Authentizität von Inhalten. Generative KI kann täuschend echte Fälschungen erzeugen und so Desinformation begünstigen. Das stellt die Glaubwürdigkeit von Medien vor Herausforderungen, untergräbt Vertrauen in Institutionen – und gefährdet letztlich die Demokratie.

Damit generative KI verantwortungsvoll genutzt werden kann, braucht es klare Regeln, ethische Leitplanken, transparente Nutzung und ein Bewusstsein für soziale Folgen. Der Blick auf mögliche Zukunftsszenarien hilft, frühzeitig zu gestalten, statt nur zu reagieren. Nur durch Zusammenarbeit aller gesellschaftlichen Akteure kann generative KI zum Wohle aller eingesetzt werden. Dies ist ein fortlaufender Aushandlungsprozess, der auch uns als Einzelne fordert!

Auch in Zukunft verspricht generative KI zahlreiche neue Möglichkeiten, die unsere Welt verändern werden. Ihre Entwicklung und Nutzung ist eine gesamtgesellschaftliche Aufgabe, die technische, rechtliche und ethische Aspekte verbindet. Entscheidend ist, ob wir als Gesellschaft die Kompetenz und den Willen aufbringen, diese Technologie aktiv und verantwortungsvoll mitzugestalten. Die Broschüre zeigt, wie ein werteorientierter Umgang mit generativer KI gelingen kann – mit Orientierung, Handlungsperspektiven und Denkanstößen.

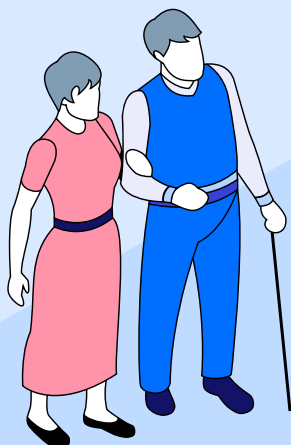
# Inhalt

<b>Wo steht die gesellschaftliche Diskussion?</b>	<b>4</b>
<b>Status quo</b>	<b>6</b>
Werteorientierte Entwicklung	7
<b>INFO</b> Generative KI im Einsatz	8
<b>Die Zukunft mit generativer KI</b>	<b>10</b>
Spannungsfelder	10
<b>INFO</b> KI-Assistenten der Zukunft	12
<b>INFO</b> Ein guter Prompt!	13
<b>Gestaltung und Partizipation</b>	<b>14</b>
Anforderungen beim Einsatz	15
<b>Handlungsansätze für eine wertorientierte Nutzung</b>	<b>16</b>
Regulierung und Ethik	18
Forschung und Entwicklung	20
Information und Medien	23
Bildung, Qualifikation und Kompetenzen	26
<b>Ausblick</b>	<b>28</b>
Generative KI: menschenzentriert und gemeinwohlorientiert	30
<b>Anhang</b>	<b>32</b>

# Wo steht die gesellschaftliche Diskussion?

**Welche Chancen und Risiken ergeben sich durch generative KI? Wie wird generative KI derzeit gestaltet? Welche Auswirkungen auf die Gesellschaft sind zu erwarten?**

Generative KI ist aktuell in der gesellschaftlichen Debatte sehr präsent, da mit der Technologie viele Erwartungen – positive wie negative – verbunden und erste Auswirkungen bereits sichtbar sind. Denn sie greift intensiv in die Art und Weise ein, wie wir leben und arbeiten. Während in der Vergangenheit Inhalte – Texte, Fotos, Videos, Bilder etc. – zumeist von Menschen erzeugt wurden und KI als Werkzeug zur Automatisierung spezifischer Aufgaben eingesetzt wurde, beispielsweise bei der Websuche, der Produktempfehlung oder der Analyse von großen Datenmengen, zeigt generative KI nicht nur eine viel größere Bandbreite, sondern auch eine menschenähnliche kreative Vielseitigkeit im Erzeugen von Text, Bild, Audio und Video. Als Sprachexpertin, in Suchmaschinen oder in Form eines intelligenten Assistenten ist sie bereits in unserem Alltag angekommen.





# Status quo

Die Entwicklung großer Sprachmodelle, die eine Grundlage von generativer KI bilden, hat in den letzten Jahren große Fortschritte gemacht. Die von den Herstellern mit generativer KI verbundenen Versprechungen scheinen nahezu allumfassend: Steigerung der Produktivität, Fortschritte im Medizinbereich, Unterstützung bei der Lösung globaler Probleme wie dem Klimawandel, Bewältigung von Pandemien und Optimierung von Lieferketten, aber auch Erleichterungen im täglichen Leben.

Mit dem Einsatz von generativer KI gehen aber auch Herausforderungen und Gefahren einher, wie etwa Cyber-Kriegsführung, Fake News oder Kontrollverlust über autonome KI-Systeme. Letztlich kann dies, auch bedingt durch eine weitgehend unregulierte Entwicklungsrichtung und Machtkonzentration der Digitalkonzerne, zu einer potenziellen Gefährdung für freiheitliche demokratische Gesellschaften und Demokratien führen. Diesen Herausforderungen gilt es, mit geeigneten Handlungsansätzen und gezielten Maßnahmen zu begegnen.

Eine zentrale Richtungsentscheidung besteht hierbei vor allem in der Frage, wie die Aushandlung der zentralen Werte **Sicherheit** und **Fortschritt** bei der Entwicklung und Gestaltung generativer KI gelingen kann und wer an diesem Prozess teilhaben sollte, also **Partizipation**. Diese Werte werden im Kontext von KI-Entwicklung oft als „Entweder-oder-Optionen“ dargestellt, was irreführend ist und zu falschen Priorisierungen führen kann. Stattdessen sollten Maßnahmen angestrebt werden, die diese miteinander verbinden.



## Generative KI

... bezeichnet KI-Systeme, die eigenständig neue Inhalte wie Texte, Bilder, Audio, Videos oder Code erzeugen. Sie nutzen große Modelle, die Wahrscheinlichkeitsverteilungen über mögliche Ausgaben berechnen und daraus neue Ergebnisse generieren. Dabei verarbeiten, übersetzen und kombinieren sie Informationen, um menschliche Arbeit zu ergänzen und komplexe Aufgaben zu automatisieren. Ihre Fähigkeit, Inhalte auf Basis erkannter Strukturen neu zusammenzusetzen, macht sie vielseitig einsetzbar.

### Übersicht über gängige Modelle:

#### Sprachmodelle:

- ChatGPT (Open AI, USA)
- Claude (Anthropic, USA)
- DeepSeek (China)
- Gemini (Google, USA)
- Llama (Meta, USA)
- Le Chat (Mistral AI, Frankreich)

#### Bild- und Multimodal:

- Stable Diffusion (Stability AI, UK/Deutschland)
- Adobe Firefly (Adobe, USA)

#### Wissenschaftliche Anwendungen:

- AlphaFold (DeepMind, UK/USA, Proteinstrukturvorhersage)
- GNoME (DeepMind, UK/USA, Materialwissenschaft, Kristallstrukturen)



GENERATIVE KI – ALLESKÖNNER DER KÜNSTLICHEN INTELLIGENZ?

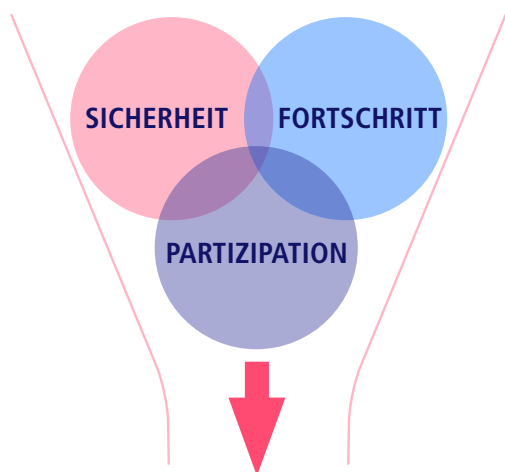


FUNKTIONSWEISE VON  
GENERATIVER KI



Um eine werteorientierte Entwicklung von generativer KI zu unterstützen, müssen die teils widerstreitenden Interessen und Erwartungen, die aus Fortschritt, Sicherheit und Partizipation abgeleitet werden, gesellschaftlich verhandelt und bewertet werden.

## Werteorientierte Entwicklung



WERTEORIENTIERTE ENTWICKLUNG/  
GESTALTUNG VON KI



**Sicherheit** bedeutet bei der Entwicklung von generativer KI, Risiken wie Missbrauch, Fehlfunktionen und Diskriminierung aktiv zu minimieren und persönliche Daten zu schützen. Dies erfordert robuste Sicherheitsmechanismen, transparente Prozesse und regelmäßige Prüfungen zur Vorbeugung potenzieller Schäden für Individuen und Gesellschaft.





**Fortschritt** bedeutet, das transformative Potenzial generativer KI im Unternehmensumfeld und zur Lösung globaler Herausforderungen zu nutzen. Dabei steht die Förderung von Innovation und technologischer Entwicklung im Zentrum, um eine bessere Zukunft für die Gesellschaft zu gestalten.

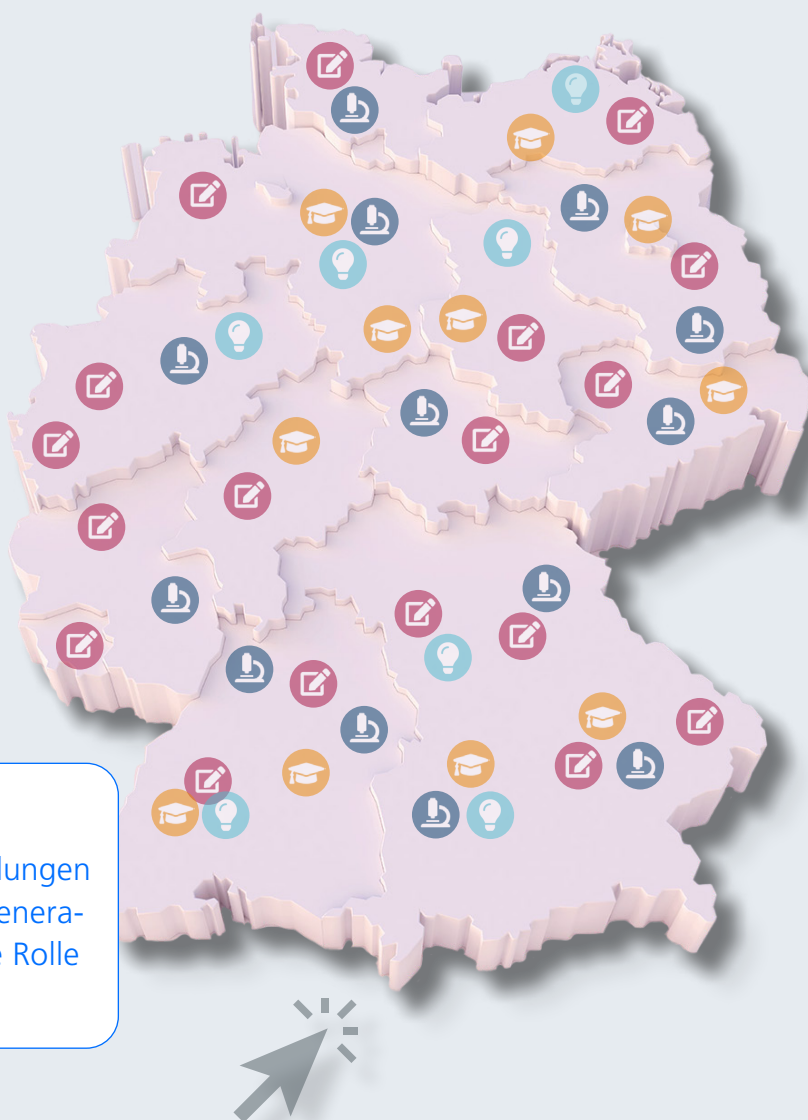
**Partizipation** bedeutet, dass die Entwicklung und Nutzung generativer KI inklusiv gestaltet wird und möglichst viele betroffene gesellschaftliche Akteure einbezieht. Dies fördert Transparenz, stärkt das Vertrauen und stellt sicher, dass unterschiedliche Perspektiven und Bedürfnisse berücksichtigt werden.



## Generative KI im Einsatz

Generative KI verändert bereits heute die Art und Weise, wie in verschiedenen Branchen gearbeitet wird. Anders als klassische KI, die vor allem Daten analysiert oder Entscheidungen unterstützt, erzeugt generative KI eigenständig neue Inhalte. In der Medienbranche schreiben KI-Systeme erste Nachrichtenentwürfe zu komplexen Themen, erstellen automatische Zusammenfassungen oder generieren multimediale Inhalte für unterschiedliche Zielgruppen. In der Industrie unterstützt generative KI im Dokumentations- und Berichtswesen oder beim Kundenservice. In der Wissenschaft wird sie zu Recherchezwecken genutzt, dient als Sparringspartner beim Brainstorming oder unterstützt bei der Dokumentation von Forschungsergebnissen. Generative KI ist damit weit mehr als ein Analysewerkzeug – sie wird zum kreativen und kollaborativen Partner in Forschung, Wirtschaft und Alltag.

-  Anwendungen
-  Forschungsinstitutionen
-  Politische Strategien & Transferzentren
-  Studiengänge



### KI-Landkarte

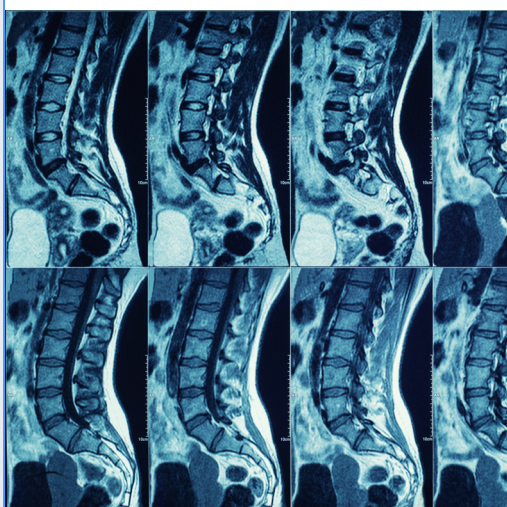
Entdecken Sie innovative Anwendungen und Projekte, bei denen KI und generative KI heute schon eine tragende Rolle spielen.



## Best Practices

### EXKURS

Teilweise werden auch andere, spezifische KI-Anwendungen mit dem Begriff „generative KI“ bezeichnet. Diese Anwendungen werden jedoch nicht durch große Sprachmodelle oder Bildgeneratoren ermöglicht, sondern durch maßgeschneiderte KI-Anwendungen, die zudem relevantes Domänenwissen explizit berücksichtigen (im Gegensatz zum rein statistischen Ansatz von großen Sprachmodellen). Solche KI-Anwendungen bieten u. a. großes Potenzial in der Medizin oder den Materialwissenschaften.

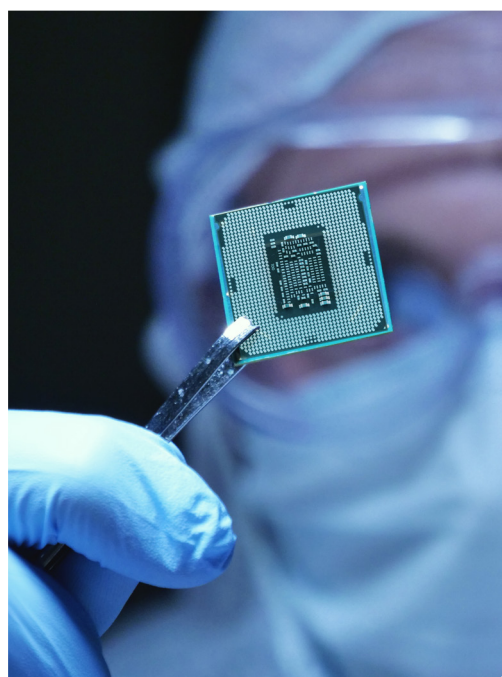


### Bildererkennung in der Medizindiagnose

In der medizinischen Bildgebung erzeugt KI synthetische Bilder zur Verbesserung von Diagnosemodellen, optimiert die Bildqualität und ermöglicht präzisere Erkennung von Anomalien. Funktionen wie Bild-zu-Bild-Übersetzungen (z. B. CT zu MRT) und realistische Simulationen komplexer Strukturen fördern personalisierte Medizin, chirurgische Planung und Forschung. KI steigert so Effizienz, Genauigkeit und Zugänglichkeit in der Diagnostik.

### Materialwissenschaften

Mit dem Tool GNoME hat die Firma Google DeepMind über zwei Millionen neue Kristallstrukturen entdeckt – darunter 380.000 stabile Materialien, die für Technologien wie Mikrochips oder Solarpanels relevant sein könnten. KI wird somit zu einem wichtigen Werkzeug in der Materialforschung und verkürzt die Entwicklungszeiten teils deutlich.



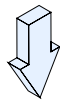
WHITEPAPER:  
GROSSE SPRACHMODELLE

# Die Zukunft mit generativer KI

Generative KI ebnet den Weg für eine Zukunft voller Möglichkeiten und Innovationen. Ihre Verbreitung und ihr Aufkommen werden von vielen utopischen, aber auch dystopischen Narrativen begleitet. Für die Gestaltung von Zukunftsszenarien auf Basis einer werteorientierten Nutzung ist es daher hilfreich, die sich herauskristallisierenden Spannungsfelder zu identifizieren und nicht die – meist ohnehin spekulativen – Eintrittswahrscheinlichkeiten verschiedener konkreter Zukunftsvisionen zu diskutieren.

## Spannungsfelder

Spannungsfelder sind meist nicht auflösbar, sehr wohl aber gestaltbar. Im Zusammenspiel mit generativer KI zeichnen sich vier Spannungsfelder ab, die aufeinander einwirken: Macht, Information und Kommunikation, Arbeit, Nachhaltigkeit. Um die darin enthaltenen kreativen Potenziale aktiv zu nutzen, ist grundsätzlich ein Bewusstsein für diese Spannungsfelder zu entwickeln. Ihre Dynamiken können dazu genutzt werden, konstruktiv auf den gemeinsamen Veränderungsprozess einzuwirken und diesen immer wieder dynamisch auszubalancieren. Denn die Auseinandersetzung mit ihnen ermöglicht vorausschauendes Handeln, gibt gesellschaftlichen Akteuren Orientierung und schafft eine Grundlage für den gesellschaftlichen Diskurs.

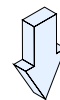


### Spannungsfeld – Generative KI und

#### **MACHT**

Es gibt Befürchtungen, dass generative KI-Anwendungen nicht-beherrschbare Risiken hervorbringen. Umso bedeutsamer ist die Frage, wer über ihre Entwicklung, den Zugang und die Einsatzbedingungen entscheidet. Derzeit liegt die Kontrolle über zentrale Technologien, Daten und Rechenkapazitäten vor allem bei wenigen großen Tech-Unternehmen, insbesondere in den USA und China. Diese verfügen über erhebliche Gestaltungsmacht – auch weil Regulierung häufig erst im Nachhinein greift, wenn überhaupt. So verlagert sich Verantwortung zunehmend auf private Akteure, die wirtschaftliche Interessen mit Verantwortung in Einklang bringen sollen. Gleichzeitig fordern

unabhängige Forschende stärkere Mitsprache – stoßen aber auf strukturelle Machtasymmetrien und begrenzten Zugang.



### Spannungsfeld – Generative KI und

#### **INFORMATION & KOMMUNIKATION**

Generative KI verändert grundlegend, wie Informationen erstellt, verbreitet und konsumiert werden. Sie ermöglicht adressatengerechte Kommunikation, unterstützt bei der Content-Erstellung und hilft, ein breiteres Publikum mit relevanten Inhalten zu erreichen – mit dem Potenzial, eine informierte Gesellschaft zu fördern. Gleichzeitig können von generativer KI erstellte Inhalte plausibel sein, jedoch inhaltlich erfunden oder völlig

falsch (sogenannte Halluzinationen). Darüber hinaus erleichtert generative KI die Erstellung und Verbreitung von Desinformation, Deepfakes oder personalisierten Inhalten in Echokammern. Dies kann die Meinungsvielfalt einschränken, das Vertrauen in Medien und öffentliche Kommunikation untergraben sowie den gesellschaftlichen Zusammenhalt und demokratische Prozesse gefährden. Erschwerend kommt hinzu, dass KI-generierte Avatare oder virtuelle Persönlichkeiten die Unterscheidung zwischen realen und künstlichen Inhalten zunehmend verwischen – was die Glaubwürdigkeit von Informationen gefährden kann.<sup>1</sup>



### Spannungsfeld – Generative KI und

#### ARBEIT

Industrialisierung und Automatisierung haben historisch selten zu echter Entlastung geführt – vielmehr zu höherem Tempo und neuen Anforderungen. Mit dem Einsatz von generativer KI stellt sich erneut die Frage: Führt sie zu mehr Freiheit und sinnvoller Arbeit oder zu weiterer Verdichtung und dem Wegfall ganzer Tätigkeiten? Die Auswirkungen auf Arbeitsprozesse, Berufsbilder und Qualifikationsanforderungen sind tiefgreifend. Um Chancen wie Entlastung durch Automatisierung oder die Bewältigung des Fachkräftemangels zu nutzen, braucht es klare Konzepte für Bildung, Kompetenzen und Teilhabe.

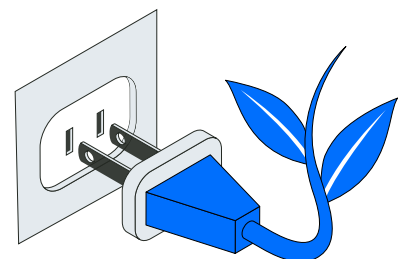
Zugleich darf nicht übersehen werden, dass viele generative KI-Systeme auf prekären Arbeitsbedingungen<sup>2</sup> und der umfassenden Verwendung geistigen Eigentums ohne Kompensation basieren. Die Gestaltung der Arbeitswelt mit generativer KI erfordert daher einen bewussten Umgang mit Ambivalenzen. Entscheidend ist: Der Mensch muss im Zentrum der Entwicklung und Anwendung stehen – nicht die Technologie.



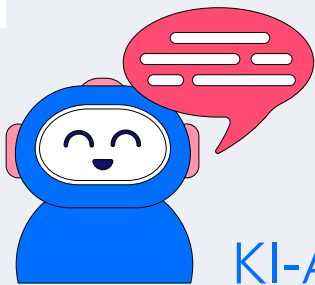
### Spannungsfeld – Generative KI und

#### NACHHALTIGKEIT

Generative KI erfordert nach wie vor enorme Rechenleistung, was erhebliche Auswirkungen auf den Energieverbrauch und die Umwelt hat. Derzeit liegt der Betrieb der leistungsfähigsten generativen KI-Modelle bei großen Technologieunternehmen, die energieintensive Rechenzentren betreiben – eine direkte regulatorische Kontrolle über deren Ressourcenverbrauch existiert bislang nicht. Technologischer Fortschritt und Verantwortung für ökologische Folgen stehen somit in einem Spannungsfeld. Einerseits besteht ein wirtschaftliches Interesse an immer leistungsfähigeren KI-Modellen, andererseits wächst der Druck, den ökologischen Fußabdruck zu reduzieren. Dazu braucht es effizientere Modellarchitekturen, den Einsatz erneuerbarer Energien und klare Transparenzvorgaben zum Energieverbrauch, um das Potenzial von generativer KI mit den Anforderungen einer nachhaltigen Zukunft in Einklang zu bringen.



<sup>1</sup> Vgl. dazu auch <https://www.die-medienanstalten.de/service/gutachten/kuenstliche-intelligenz/>  
<sup>2</sup> In sogenannter KI-Klickarbeit bereiten Menschen – teils in stundenlanger repetitiver Arbeit – Daten durch Kategorisierung und Datenbereinigung für das Training von KI-Modellen vor.

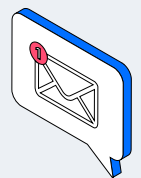


## KI-Assistenten der Zukunft

Digitale Assistenten (KI-Agenten) zählen zu den Anwendungen von generativer KI, denen aktuell das größte Potenzial zugetraut wird. Hochrangige Entwickler auf dem Gebiet der generativen KI erwarten, dass bereits in fünf Jahren jeder Zugang zu einem persönlichen Bot oder „Chief of Staff“ haben kann, der als Coach oder persönlicher Lebensbegleiter im Alltag fungiert. Zukünftige Modelle könnten über einfache Informationsabfragen hinausgehen – etwa durch Kombination von Spracheingabe, visuellem Verständnis und direkter Interaktion. Sie könnten Aufgaben wie die Zusammenfassung oder Aufbereitung von Texten, Terminorganisation oder Zahlungsabwicklung übernehmen – im privaten wie beruflichen Kontext.

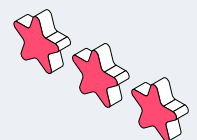
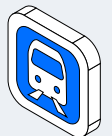
### Digitale Assistenten im Unternehmen

Im Unternehmensumfeld wirken sie als „Copiloten“: Sie automatisieren Kundenkommunikation, priorisieren E-Mails, erstellen Briefings und Protokolle, koordinieren Prozesse und senken insgesamt die Kosten. Perspektivisch könnten sie eigenständig Projekte verwalten oder Datenanalysen durchführen.



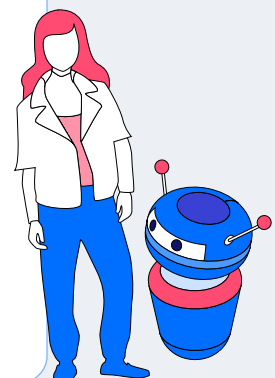
### Digitale Assistenten im Alltag

Auch im Alltag werden KI-Assistenten zentrale Funktionen übernehmen. Statt mehrere Apps zu bedienen, könnte künftig eine sprachgesteuerte Interaktion mit einer zentralen Anwendung genügen, die dann für Reiseplanung oder Haushaltsorganisation genutzt werden könnte. Durch Integration in Smart Devices wie Brillen oder Haushaltsgeräte könnte KI zur nahtlosen Schnittstelle zwischen Nutzerin/Nutzer und digitaler Welt werden.



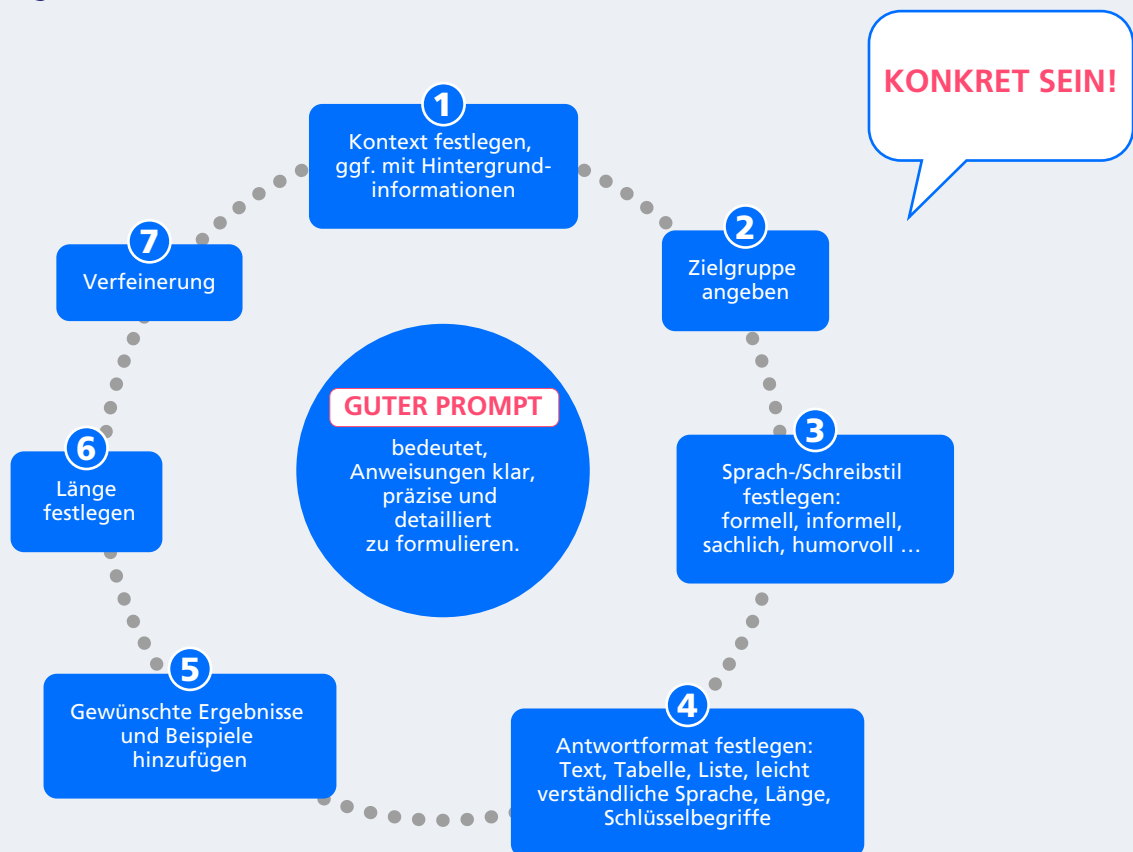
*Generative KI wird die Art und Weise, wie Menschen mit Computersystemen interagieren, komplett verändern. In ein paar Jahren wird niemand mehr Formeln für Excelfunktionen suchen, sondern dem KI-Agenten sagen, was für eine gewünschte Ausgabe man haben will.“*

**Prof. Dr. Björn Eskofier**, Friedrich-Alexander-Universität  
Erlangen-Nürnberg



# Ein guter Prompt!

Prompting ist eine Eingabeaufforderung an ein Sprachmodell. Je nachdem, wie allgemein oder wie spezifisch ein Prompt gestellt wird, fällt die Antwort generativer KI-Modelle unterschiedlich aus. Prompts können in verschiedenen Formaten vorliegen, beispielsweise als Fragen, Anweisungen oder Beispiele. Je präziser und klarer die Eingabeaufforderung, desto besser das Ergebnis!



Die Aufgabe in kleine Schritte zerlegen:

- **Ziel festlegen:** Was will ich vom KI-Tool? Informationen, Text, Ideen oder Erklärungen?  
**Beispiel:** Fasse den vorliegenden Text in Form von Aufzählung der wichtigsten Punkte zusammen.
- **Kontext geben:** Hintergrundinformationen und Beispiele liefern: Zielgruppe, Stil (Ton, Schreibstil freundlich, witzig, professionell)  
**Beispiel:** Schreibe einen Text zum Thema generative KI in leichter und verständlicher Sprache, der Jugendliche anspricht.

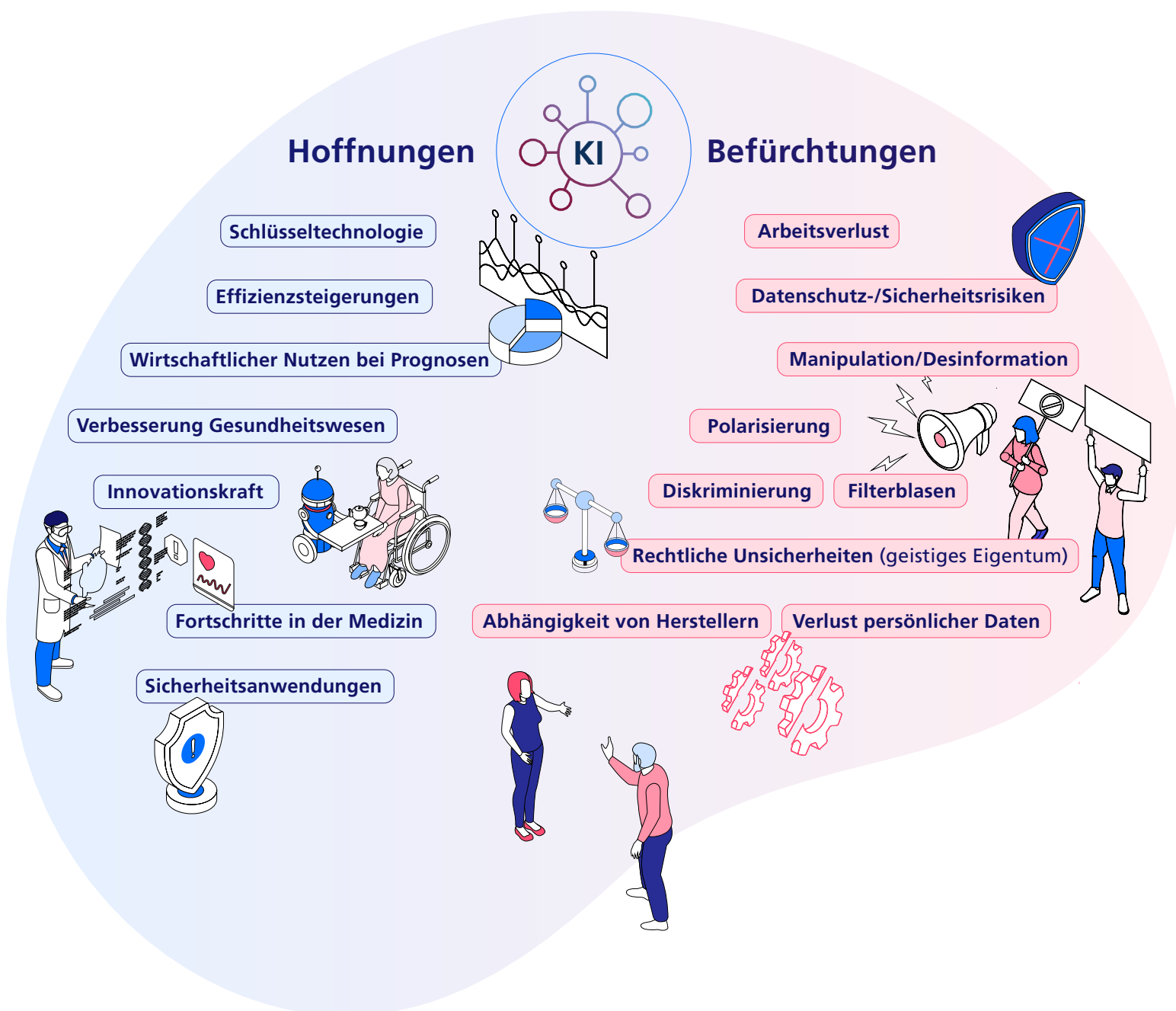
- **Format vorgeben:** Wie will ich die Antwort haben? Liste, Fließtext, Tabelle, Schritt-für-Schritt-Anleitung?  
**Beispiel:** Gib die Antworten humorvoll und witzig in Form eines kurzen Textes, drei bis fünf Sätze, aus.
- **Nachbessern & Verfeinern:** Wenn die Antwort nicht zufriedenstellend ist, nachfassen, d. h. nachfragen, ergänzen oder klarstellen und sich weitere Vorschläge vorgeben lassen.  
**Beispiel:** Kannst du die Antwort kürzer fassen? oder Gib fünf verschiedene Varianten zu diesem Text aus!



# Gestaltung und Partizipation

Vor dem Hintergrund der vier Spannungsfelder und mit Blick auf die wünschenswerten beziehungsweise zu vermeidenden Zukunftsszenarien ist zu klären, welcher aktive Handlungsspielraum den jeweiligen Beteiligten zukommt, damit alle Erwartungen und Perspektiven in die Entwicklung und Gestaltung dieser Technologie einfließen können. Dies erfordert einen demokratischen Aushandlungsprozess sowie einen Konsens über Ziele,

Möglichkeiten und Instrumente für eine zukunftsfähige Technikgestaltung. Dies setzt voraus, dass die Gesellschaft, beispielsweise in Form von Vertreterinnen und Vertretern aus Wissenschaft, Unternehmen und der Zivilgesellschaft, gemeinsam Strategien für ein gutes Mensch-Technik-Verhältnis beim Einsatz von generativer KI entwickeln muss – also eine Technikgestaltung, bei der der Mensch im Zentrum steht.





*„Menschliche Expertise und die Fähigkeit zur kritischen Beurteilung und Korrektur von generierten Inhalten ist unverzichtbar. Wir müssen darauf achten, dass generative KI und menschliche Expertise sinnvoll in einem Co-Creations-Prozess zusammenspielen, um die Zuverlässigkeit und Qualität von generierten Inhalten sicherzustellen.“*

**Prof. Dr. Ute Schmid**, Otto-Friedrich-Universität Bamberg

## Anforderungen beim Einsatz

Bei der Nutzung von generativer KI gibt es gerade auch zu Fragen hinsichtlich Regulierung oder Kompetenzentwicklung zahlreiche Überschneidungen mit allgemeinen KI-Themen. Denn durch die starke Verbreitung von generativer KI (insbesondere ausgelöst durch große Sprachmodelle) ist der Bedarf an entsprechenden Maßnahmen deutlich gestiegen.

Eine verantwortungsvolle Nutzung von KI, insbesondere generativer KI, die den ethischen, rechtlichen und praktischen Rahmenbedingungen entspricht, heißt für:

### Nutzende

- sich ein grundlegendes Verständnis aneignen, was KI ist und wie sie funktioniert
- KI verantwortungsbewusst nutzen (sich ethischer Fragestellungen wie Bias und Fairness bewusst sein)
- eigene personenbezogene und vertrauliche Daten sparsam einsetzen
- KI-generierte Inhalte überprüfen und kritisch bewerten
- KI gegenüber Dritten transparent nutzen
- KI-Anwendungen seriöser Hersteller nutzen



### Generative KI und Gute Arbeit

Um die Potenziale von generativer KI für Gute Arbeit und Produktivität zu nutzen, sind mehrere Aspekte zu berücksichtigen. Wichtig ist vor allem, generative KI immer kontextbezogen zu betrachten und vor der Einführung risikobasiert zu differenzieren. Da sich die Auswirkungen im realen Betrieb entscheiden, braucht es eine Unternehmenskultur, in der die Beschäftigten aktiv und informiert über den Einsatz von KI mitentscheiden können. Als erfolgversprechend hat sich dafür eine Art „KI-Instanz“ erwiesen: Dabei steuern die relevanten und tangierten Akteursgruppen den Prozess gemeinsam.

### Unternehmen und Organisationen

- Einsatzbereich und Einsatzzweck werteorientiert aufstellen und verbindlich festlegen
- klare und verbindliche Regelungen zur Nutzung von KI vorgeben
- geeignete KI-Systeme auswählen und Sozialpartner bei Einführung einbeziehen
- datenschutzkonformen Umgang bei der Datenverarbeitung sicherstellen
- Kompetenzaufbau seitens der Mitarbeitenden gewährleisten (regelmäßige Schulungen, Fehlertoleranz)
- Transparenz über die Nutzung von KI intern und extern offen kommunizieren (Selbstverpflichtung, Kennzeichnung)

# Handlungsansätze für eine werte- orientierte Nutzung

**Was sind die Voraussetzungen für eine werteorientierte Entwicklung und Nutzung von generativer KI?**

**Welche Rolle nehmen staatliche Institutionen, gesellschaftliche Akteure sowie wir als Einzelne dabei ein?**

Für eine werteorientierte Nutzung von generativer KI – wie auch für KI im Allgemeinen – sind verschiedene Dimensionen entscheidend. So müssen staatliche Institutionen und Behörden durch den eigenen Einsatz der Technologie Vorbild sein und klare Richtlinien setzen, um Missbrauch zu verhindern und gleichzeitig ethische und rechtliche Standards zu wahren. Forschung und Entwicklung sind aufgerufen, diese Standards in Innovationen und wissenschaftliche Ergebnisse zu integrieren. Was die (fehlende) Authentizität generativ erstellter Inhalte betrifft, bedarf es einer verlässlichen Kennzeichnung. Dies betrifft allen voran die Medien. Daneben ist auch Bildung wichtig, um die Funktionsweise und Auswirkungen von generativer KI verstehen und fundierte Entscheidungen treffen zu können. Nur eine KI-mündige Gesellschaft kann KI-Entwicklungen und ihre Ergebnisse überprüfen und kritisch begleiten. Erst im Zusammenspiel aller Akteure und den damit verbundenen Dimensionen kann generative KI zum Wohle aller gestaltet werden. Aus rechtlicher Sicht besonders relevant ist hierzu der AI Act (KI-Verordnung), der Anforderungen für KI-Anwendungen definiert, darunter auch die Kennzeichnung von Inhalten und die Entwicklung von KI-Kompetenzen für Beschäftigte.





# Regulierung und Ethik

In der Debatte um generative KI sind Regulierung und Ethik keine Gegensätze, sondern ergänzen sich: Während Regulierung durch rechtliche Vorgaben konkrete Handlungsrahmen und damit Rechtssicherheit schafft, sorgt Ethik dafür, dass diese Handlungsrahmen auf Grundwerten wie Gerechtigkeit, Transparenz und Verantwortung beruhen. Diese Kombination aus klaren Regeln und einer wertebasierten Perspektive schafft im Idealfall eine Balance zwischen Sicherheit, Innovation und individueller Verantwortung, die immer wieder neu angepasst und ausgehandelt werden muss.

## REGULIERUNG konkreter Handlungsrahmen

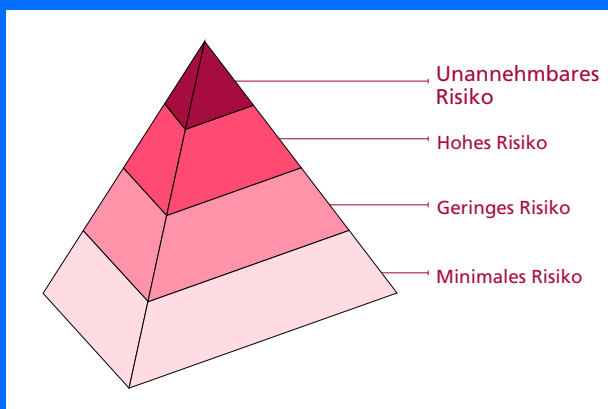
Regulierung dient als strukturelles Instrument, indem sie Grenzen und Standards definiert, die sicherstellen, dass KI-Systeme sicher, transparent und diskriminierungsfrei agieren. Wenn Regulierung gut gelingt, schafft sie einen verlässlichen rechtlichen und organisatorischen Rahmen für die Entwicklung und Nutzung von (generativer) KI und kann Unternehmen ermutigen, aktiv in deren Entwicklung und Nutzung zu investieren.

Im Kontext des AI Act bedeutet dies die Einführung klarer Vorgaben für Anbieter und Betreiber, um Missbrauch und Schäden für die Gesellschaft zu verhindern und zugleich Innovationen zu fördern: Regulierung ist also nicht als Hemmnis für Innovation, sondern vielmehr Treiber für verantwortungsvollen Fortschritt zu verstehen.



### Artificial Intelligence Act der EU (kurz: AI Act)

Die EU-Verordnung zum AI Act, die erste umfassende Gesetzgebung der EU zur Regulierung von KI, wurde 2024 verabschiedet. Die Verordnung verfolgt einen risikobasierten Ansatz in vier Stufen: Sie kategorisiert KI-Anwendungen nach ihrem Gefährdungspotenzial, von minimalem bis hin zu inakzeptablem Risiko. Seit dem Aufkommen von generativer KI mit ChatGPT im November 2022 wurde deutlich, dass diese Technologie besondere Herausforderungen mit sich bringt, beispielsweise hinsichtlich Transparenz, Urheberrecht, Verantwortung und (IT-)Sicherheit. Daher wurden im AI Act spezifische Regelungen für leistungsfähige generative KI-Anwendungen (unter der Bezeichnung General-Purpose AI, GPAI) aufgenommen, die im Juli 2025 durch den GPAI Code of Practice in konkrete Vorgaben überführt wurden. Dieser bietet betroffenen Unternehmen eine praxisnahe und rechts-sichere Umsetzung der Regelung für GPAI-Systeme.



LESETIPP:  
KI KOMPAKT: AI ACT DER EUROPÄISCHEN UNION



ARTIFICIAL INTELLIGENCE ACT DER EU



*„Regulierung darf nicht auf Deutschland und Europa beschränkt bleiben, sondern es sollte eine internationale ‘Aufsichtsbehörde’ geben – vielleicht unter dem Dach der UN, ähnlich den Strukturen der Internationalen Atomenergiebehörde IAEA. Erste Anstrengungen in diese Richtung werden bereits unternommen.“*

**Prof. Dr. Björn Eskofier**, Friedrich-Alexander-Universität Erlangen-Nürnberg

## ETHIK

### Fundament für individuelles und gesellschaftliches Handeln

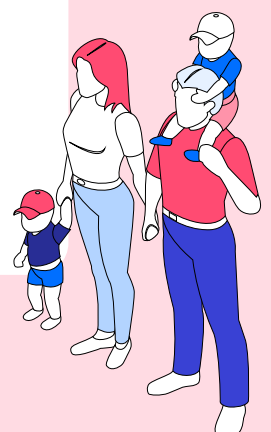
Ethik bildet die Grundlage für die Regulierung und stellt Fragen nach den Werten und Prinzipien, die den Einsatz von KI im Allgemeinen und generativer KI leiten. Sie fordert eine individuelle und gesellschaftliche Auseinandersetzung: Wie möchte ich KI nutzen und welche Verantwortung trage ich?

Sozialethisch relevant sind vor allem Gerechtigkeit, Teilhabe und Solidarität. Denn der Einsatz von generativen KI-Technologien kann soziale Ungleichheiten verstärken oder abbauen – je nachdem, wie sie entwickelt und eingesetzt werden. Daher müssen Chancengleichheit, Inklusion und der Schutz vor Diskriminierung zentrale Leitprinzipien sein. Zudem beeinflusst generative KI die Arbeitswelt, Bildung und soziale Sicherungs-

systeme. Automatisierung schafft neue Möglichkeiten, kann aber auch Arbeitsplätze verdrängen. Die Gestaltung dieses Wandels erfordert Weiterbildung und faire Übergänge. Partizipation ist ein wichtiger Faktor: Entscheidungen über KI dürfen nicht nur wenigen Akteuren überlassen werden, sondern müssen demokratisch und nicht von einzelnen Akteuren verhandelt werden. Mit dem globalen Einsatz von generativer KI geht die Verantwortung für ökologische Nachhaltigkeit und gerechten Zugang zu technologischen Ressourcen einher.



Für die Unternehmen wie die Gesellschaft bedeutet dies, nicht nur rechtliche Vorgaben zu beachten, sondern sich auch aktiv mit der Frage auseinanderzusetzen, welche Ziele und Grenzen bei der Entwicklung und für den Einsatz von generativer KI gesetzt werden sollten.

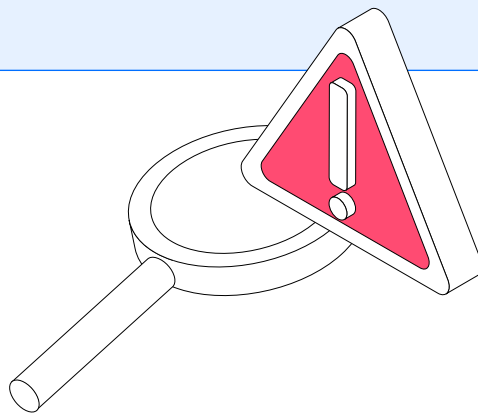


# Forschung und Entwicklung

Forschung und Entwicklung sollten sich nicht nur mit der rein technischen Weiterentwicklung befassen, sondern gleichzeitig auch mit einer wertorientierten Gestaltung von generativer KI, die den ethischen und regulatorischen gesellschaftlich anerkannten Anforderungen und Werten gerecht wird. Dies fördert und stärkt das Vertrauen in die Leistungsfähigkeit, Zuverlässigkeit und Sicherheit generativer KI-Systeme – insbesondere in sensiblen Anwendungsbereichen wie der Medizin. So kann der Grundstein für eine vertrauenswürdige generative KI gelegt werden. Ein besonderes Augenmerk ist dabei auf Halluzinationen zu legen, ein inhärentes Problem großer generativer Sprachmodelle. Hier müssen Mechanismen geschaffen werden, die solche Fehler minimieren und deren Erkennung erleichtern.

## Kriterien und Prüfungsansätze für vertrauenswürdige (trustworthy) generative KI:

- **Transparenz:** Hinweis auf die Beteiligung von generativer KI, verwendete Datenbasis etc.
- **Erklärbarkeit:** Verständlichkeit der Funktionsweise und Entscheidungsprozesse
- **Nachvollziehbarkeit:** Dokumentation und Überprüfung der Prozesse und Entscheidungen
- **Fairness:** Vermeidung von Diskriminierung und Bias in den Ergebnissen
- **Sicherheit:** Schutz vor Fehlfunktionen, Missbrauch und externen Angriffen
- **Datenschutz:** Sicherstellung der Vertraulichkeit und des verantwortungsvollen Umgangs mit sensiblen Daten
- **Robustheit:** Zuverlässigkeit und Widerstandsfähigkeit unter verschiedensten Bedingungen





## HALLUZINATIONEN

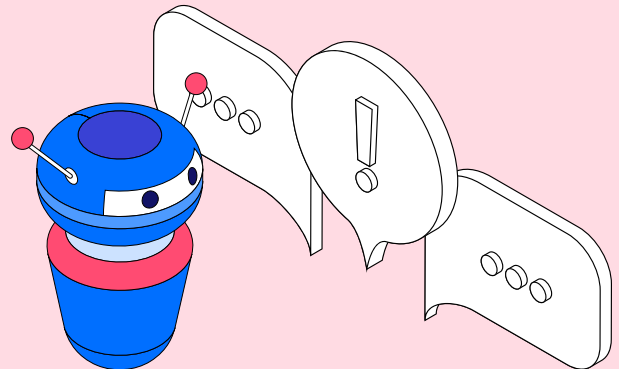
Bisweilen erzeugen generative Systeme falsche Informationen und Fakten (Halluzinieren). Große Sprachmodelle generieren diese Inhalte anhand von Mustern und Zusammenhängen. Sie nutzen dabei eine statistische Funktionsweise und haben kein Verständnis für die Bedeutung und den Inhalt der von ihnen erzeugten Texte. Solche Halluzinationen können schwerwiegende Folgen haben, wenn falsche Informationen für wahr gehalten werden, gerade in kritischen Bereichen wie der Medizin, Nachrichtenberichterstattung oder für demokratische Wahlen. Dieses Problem besteht bereits, ohne dass feindlich

gesinnte Akteure mit dem Ziel der Disruption von gesellschaftlichen Institutionen dahinter stehen!

**Beispiel:** Im Vorfeld der Landtagswahlen 2023 in Bayern und Hessen wurde berichtet und in Studien gezeigt, dass generative KI-Anwendungen falsche Informationen über die Kandidatinnen und Kandidaten, die Wahlprogramme oder die Rahmenbedingungen der Wahl (etwa Termine oder Mindestalter für Wählerinnen und Wähler) produzierten.



Seitens der Forschung sind Tools und Methoden zu entwickeln, die generative KI hinsichtlich dieser Kriterien und Prüfungsansätze bewerten und zertifizieren können – hier werden Ansätze wie werteorientierte Technikgestaltung („Ethics by Design“) mit technischen Prüfverfahren kombiniert.



[INTERVIEW MIT ALEXANDER LÖSER:  
KANN EUROPA BEI SPRACHMODELLEN AUFHOLEN?](#)



[WHITEPAPER: ZERTIFIZIERUNG  
VON KI-SYSTEMEN](#)



## Mensch- und wertezentrierte Technikgestaltung

Ansätze für eine verantwortungsvolle generative KI-Entwicklung:

- **Alignment**, das heißt ein Übereinstimmen der Ziele und Verhaltensweisen von (generativer) KI mit den Werten, Interessen und Erwartungen der Entwickelnden und Nutzenden. Hierfür sind sowohl ethische Richtlinien als auch technische Mechanismen notwendig.
- **Kennzeichnung von „KI-Decision-Inside“-Anwendungen**, die einen gewissen Autonomiegrad besitzen und KI-basiert Entscheidungen treffen. Ziel ist es, das Bewusstsein zu erhöhen, in welchen Anwendungen (generative) KI bereits eine wichtige Rolle spielt, und die oft bestehende Diskrepanz zwischen dem Umgang mit KI im beruflichen und privaten Bereich sichtbar zu machen.
- **Regelmäßige Schulung von Entwicklerinnen und Entwicklern** rund um das Thema (generative) KI zum aktiven Mitdenken von Nutzungsszenarien bei der Gestaltung und Entwicklung von Anwendungen.



*„Ganz gleich, wo KI zum Einsatz kommt, stellt sich immer die Frage nach der Vertrauenswürdigkeit. Wichtig ist deshalb, dass der Mensch stets in einer aktiven Rolle bleibt. Unternehmen stehen in der Verantwortung, die Vertrauenswürdigkeit zu stärken. Um Chancen und Herausforderungen der Technologie angemessen zu steuern, ist eine effektive KI-Governance unerlässlich. Diese umfasst Richtlinien, Standards und Verfahren, die sicherstellen, dass KI-Systeme verantwortungsvoll entwickelt und eingesetzt werden.“*

**Andrea Martin**, IBM Deutschland GmbH

# Information und Medien

Mit generativer KI lassen sich Texte, Bilder, Audios oder Videos produzieren. Den möglichen Inhalten sind dabei kaum Grenzen gesetzt: Von Beiträgen, die historische Persönlichkeiten zum Leben erwecken und so Geschichte greifbar machen, bis hin zu gezielter Desinformation mit Deepfakes von Politikerinnen und Politikern – all diese KI-generierten Inhalte haben eines gemeinsam: Sie erzeugen neue Informationen. Und mit den Möglichkeiten des Internets verbreiten sich diese zudem, oft ungeprüft, rasend schnell. Für Nutzerinnen und Nutzer wird es so immer schwieriger zu erkennen, ob ein Beitrag, ein Videointerview, ein Bild nun Fake oder Fakt ist. Ein kompetenter Umgang mit Quellen, Informationen und Nachrichten aller Art ist daher unabdingbar – dabei sollen auch der AI Act und weitere Regulierungen die Nutzerinnen und Nutzer unterstützen!

## Deepfake von John F. Kennedy: Appell in die Zukunft

Im Juni 1963 besuchte der damalige US-Präsident John F. Kennedy Deutschland und hielt in der Frankfurter Paulskirche eine Rede.

Welche Herausforderungen würde John F. Kennedy im KI-Zeitalter betonen? Forschende der TU Darmstadt haben mit generativer KI eine Antwort versucht.

Das Beispiel zeigt, wenngleich als Deepfake, welche Möglichkeiten sich für Bildung und Lehre mit generativen KI-Systemen ergeben, wenn sie historische Persönlichkeiten mit all ihren damaligen Gegebenheiten und Herausforderungen als kurzweiliges wie lehrreiches Element zum Leben erwecken – allerdings mit verpflichtender Kennzeichnung, wenn generative KI eingesetzt wird.

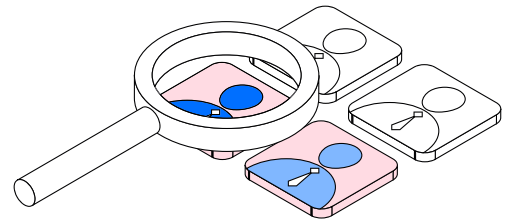


**Appell in die Zukunft** – Das System Security Lab der TU Darmstadt testete die Möglichkeiten frei zugänglicher KI-Tools für Deepfakes mit Bildungsanspruch.

Quelle: <https://youtu.be/4rhqtAOCOYs>



## Generative KI für Zugang zu mehr Informationen und Wissen



Eines der großen Potenziale von generativer KI liegt in der Möglichkeit, mehr Informationen und Wissen für einen größeren Teil der Bevölkerung zugänglich zu machen. So können mit generativer KI beispielsweise Sprachbarrieren unkompliziert überwunden sowie relevante Informationen schnell gesammelt und Antworten personalisiert

auf die Bedürfnisse und Vorkenntnisse der Nutzenden zugeschnitten werden. Bei der Bereitstellung von Informationen und Antworten müssen jedoch separate Maßnahmen zur Qualitätssicherung mitgedacht werden, um dem Problem von Halluzinationen zu begegnen.

**Fakt oder Fake? Unser Quiz zum Mitmachen** [↗](#)



### Generative KI im Journalismus

Medien sind entscheidende Akteure bei der Weitergabe von Informationen und fördern somit die öffentliche Meinungsbildung und Partizipation. Als wichtiger Bestandteil einer demokratischen Gesellschaft sortieren, präsentieren und kommentieren sie Fakten kritisch. Für die Herstellung von Nachrichten wurde schon immer auf technische Hilfsmittel zurückgegriffen. Mit generativen KI-Systemen als wirkmächtigem Werkzeug erweitern sich die Möglichkeiten, eigenständig neue Beiträge (sog. synthetische) zu schaffen: indem vorhandene Beiträge in Form von Text, Audio oder Video in eine andere Version überführt werden – beispielsweise Text zu Audio oder umgekehrt. Oder sogar völlig neue Beiträge erstellt werden, wie etwa bei der automatisierten Berichterstattung über Sportereignisse oder Börsenwerte. Solche KI-generierten Informationen sollten als vertrauensbildende Maßnahme stets gekennzeichnet, transparent und verifizierbar sein.

### Einsatzfelder von genKI in Redaktionen



© Plattform Lernende Systeme



## Auswirkungen auf die Meinungsbildung

Durch Manipulation mittels überzeugender Bildfälschungen, Falschinformationen, Deep-fakes kann der demokratische Diskurs und die individuelle Meinungsbildung – gerade im Vorfeld von Wahlen – beeinflusst werden. Auch aufgrund der Informationsflut ist oft nicht mehr verifizierbar und erkennbar, ob es sich um Fakten oder Fake handelt. Genauso können aber auch Zweifel an der Authentizi-

tät der Informationen entstehen, die das Vertrauen in wissenschaftliche oder journalistische Institutionen untergraben können: Hervorgerufen durch strategisches Platzieren gefälschter Inhalte oder durch Verdrängung der Meinungsbildung aus der Öffentlichkeit ins Private in die Interaktion mit Chatbots/Sprachmodellen.



Um Desinformation zu minimieren, gilt es Inhalte stets kritisch zu hinterfragen, Quellen auf Vertrauenswürdigkeit zu überprüfen und im Zweifelsfall fragwürdige Inhalte nicht weiterzuleiten oder zu teilen. In diesem Kontext sollten mit generativer KI erzeugte Inhalte stets als solche gekennzeichnet sein.




# Bildung, Qualifikation und Kompetenzen

Generative KI ist vielseitig einsetzbar und kann in unterschiedlichsten Arbeitskontexten Inhalte und Lösungen liefern und so die Produktivität und die Effizienz in vielen Berufen signifikant erhöhen. Das setzt voraus, dass die Nutzenden in der Lage sind, die Technologie kompetent zu bedienen und einzusetzen. Denn ein unsachgemäßer Umgang kann zu unerwünschten oder sogar schädlichen Folgen für Beschäftigte wie Arbeitgeber führen, etwa wenn Beschäftigte mit sensiblen Daten arbeiten oder mit generativer KI erzeugte Inhalte ohne Prüfung in wichtigen Angelegenheiten weiter verwenden. Daher müssen die notwendigen Kompetenzen bei den Nutzenden identifiziert und geschult werden und Technologien so gestaltet werden, dass die Hürden für eine erfolgreiche Nutzung so weit wie möglich reduziert werden.

## Voraussetzungen für einen kompetenten Umgang in Unternehmen

- **Reflexion/kritische Haltung:** Nutzende müssen sich bewusst sein, dass die produzierten Inhalte trotz der stilistisch überzeugenden Präsentation nicht notwendigerweise korrekt sind (Problem Halluzination).
- **Datenschutz:** Bei der Nutzung von generativer KI ist darauf zu achten, dass bei der Interaktion keine personenbezogenen oder sensiblen Daten des Unternehmens an die Modellhersteller abfließen oder öffentlich werden.
- **Transparenz:** Eine offene Kommunikation, die für Transparenz und Austausch steht, sollte vorherrschen, um Ängste und Hemmnisse abzubauen – als vertrauensbildende Maßnahme gegenüber Mitarbeitenden, Kunden und anderen Stakeholdern.
- **Qualitätskontrolle:** Unternehmen sollten sicherstellen, dass ihre Systeme auch nach längerem Einsatz zuverlässige und diskriminierungsfreie Ergebnisse liefern, z. B. durch Rahmenbedingungen wie Use Cases, Diskussionsrunden oder eine Experimentierumgebung (Fehlerkultur).
- **Training und Kompetenzaneignung:** Unternehmen sollten regelmäßig zielgruppengerechte Schulungen zu Möglichkeiten und Grenzen generativer KI anbieten. Diese sollten technische sowie ethische Aspekte abdecken, damit Mitarbeitende souverän damit umgehen können.



Wie der KI-Einsatz in Unternehmen gelingen kann, veranschaulicht der **Information-Butler fürs Büro**



Für eine erfolgreiche Gestaltung der Zukunft mit generativer KI muss eine Ausgewogenheit gefunden werden, zwischen dem, was die Nutzenden leisten müssen – Entwicklung von Daten- und KI-Kompetenz –, und dem, wie die Technologie entwickelt und konzipiert werden muss – nach Prinzipien wie Erklärbarkeit, Transparenz etc.

## Kompetenzentwicklung

Die Identifikation, Entwicklung und Stärkung zentraler Kompetenzen ist entscheidend – nicht nur für Mitarbeitende in Unternehmen, sondern letztlich für alle Nutzenden. Denn die zunehmende Durchdringung aller Bereiche des täglichen Lebens durch vernetzte KI-Systeme erfordert Vertrauen in und Wissen über generative KI als neuen Interaktionspartner. Auch die Kenntnis ihrer Grenzen.

Kompetenter Umgang mit generativer KI:

- **Weiterentwicklung von individuellen Kompetenzen**
  - a) **Kompetenzen, um mit generativer KI zu kommunizieren und zu interagieren:** Dazu zählt vor allem das Formulieren geeigneter Prompts, die weit über klassische Suchmaschinenanfragen hinausgehen.
  - b) **Kompetenzen, um Ergebnisse von generativer KI richtig einzuordnen und zu bewerten:** Neben Hintergrundwissen sind hierbei vor allem kritisches Denken und Urteilsvermögen gefragt.
  - c) **Kompetenzen, um zu bewerten, wann der Einsatz generativer KI sinnvoll ist:** Dazu zählen Wissen und Kenntnisse über Funktionsweise von generativer

KI, Daten und ihre Qualität, aber auch wo die Grenzen von generativer KI liegen. Das schließt die frühzeitige Wissensvermittlung in schulischer Bildung und Studium mit ein sowie die Schaffung von Experimentierräumen.

- **Soziale Kompetenzen**

**Kommunikations- und Konfliktfähigkeit, um dem zunehmenden Aufkommen von generativer KI adäquat zu begegnen:** Diese Fähigkeiten erlangen zunehmend eine stärkere Relevanz, insbesondere dann, wenn generative KI-Systeme im Team Assistenztasks übernehmen und damit mehr Freiräume für Interaktion und Zusammenarbeit zwischen den Menschen im Team entstehen.

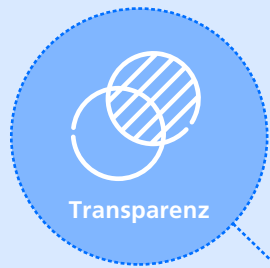
- **Technologieadaption seitens der Unternehmen**

**Generative KI-Nutzung fördern, um ihr Potenzial voll auszuschöpfen:** Der Einsatz bietet nicht nur Unternehmen, sondern auch Mitarbeitenden viele Möglichkeiten: von neuen Wertschöpfungsmodellen über Effizienzsteigerungen bis hin zur Entlastung und Unterstützung von Tätigkeiten.

# Ausblick

**Generative KI als Allzweck-KI – welche Rolle wird sie bei verschiedensten Unternehmens- und Verbraucherzwecken in Zukunft spielen? Wie werden uns diese Technologien im beruflichen und alltäglichen Leben begleiten? Welche Risiken sind damit verbunden und wie können diese abgemildert werden?**

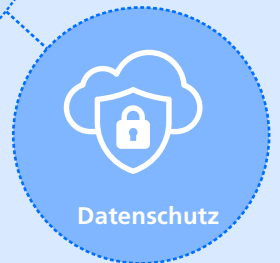
Generative KI wird als Allzweck-KI-Technologie unser tägliches Leben sowie berufliche Abläufe stark transformieren und ein steter Begleiter zur Bewältigung von vielfältigsten Aufgaben sein. Während sie viele Prozesse vereinfachen und neue kreative Wege eröffnen kann, gilt es gleichzeitig, den verantwortungsvollen und sicheren Umgang mit dieser Technologie zu fördern. Die zukünftige Zusammenarbeit zwischen Mensch und generativer KI als effektives Werkzeug, als Chatbot oder als eigener KI-Agent wird entscheidend sein – sie muss auf Vertrauen, Transparenz und kontinuierlicher Weiterbildung basieren, um die zahlreichen Vorteile nachhaltig zu nutzen.



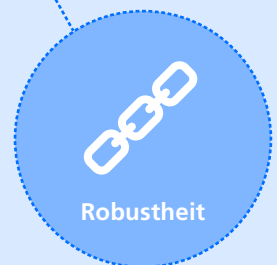
Transparenz



Erklärbarkeit



Datenschutz



Robustheit



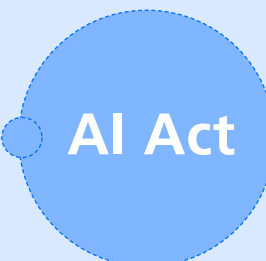
Sicherheit



Fairness



Nachvollziehbarkeit



AI Act

# Generative KI: menschzentriert und gemeinwohlorientiert

Eine erfolgreiche Verbindung von Sicherheit, Fortschritt und Partizipation in einem ausgewogenen Gleichgewicht ist entscheidend für die Entwicklung und Gestaltung von generativer KI, die sich für das Gemeinwohl positiv auswirkt. Folgende Ansätze könnten dazu beitragen, eine Balance zwischen den drei Werten herzustellen:

- **Fokus auf menschenzentrierte und gemeinwohlorientierte Anwendungen**

Anstatt ausschließlich auf immer leistungsfähigere Modelle zu setzen, sollte der Schwerpunkt auf der Entwicklung und Nutzung bestehender Technologien für konkrete gesellschaftliche Herausforderungen liegen. Dies fördert Innovation, ohne Sicherheitsrisiken unnötig zu erhöhen.

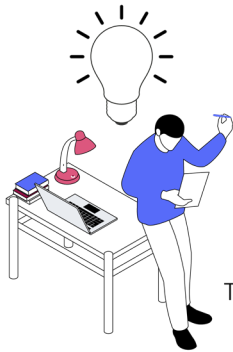
- **Forschung zu sicherer und robuster und damit kontrollierbarer generativer KI**

Eine gezielte Förderung der Forschung zu Sicherheitsmechanismen ist entscheidend, um sicherzustellen, dass generative KI-Systeme transparent, nachvollziehbar und zuverlässig sind und entsprechend den Werten und Zielsetzungen funktionieren

(Alignment). Techniken wie erklärbare KI, hybride und robuste Modelle, unterstützt durch ethische Leitlinien und regulatorische Vorgaben, werden eine wichtige Rolle dabei spielen, Risiken zu minimieren

- **Demokratische Entscheidungsfindung**

Um den Einsatz von generativer KI mit gesellschaftlichen Werten in Einklang zu bringen, sollten breite, inklusive Diskussionen stattfinden. Partizipative Prozesse, die Menschen aus unterschiedlichen gesellschaftlichen Gruppen einbeziehen, können helfen, ethische Standards und auch Einsatzgrenzen für generative KI festzulegen und diese sinnvoll in entsprechende Strukturen und Prozesse einzubinden.

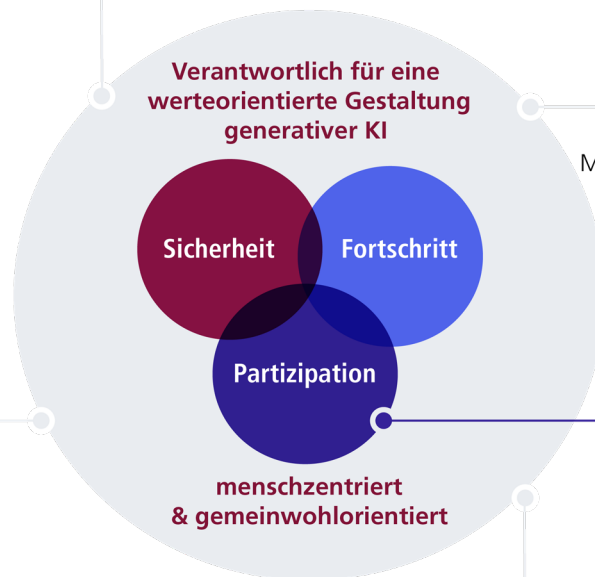


**Bei der Entwicklung  
und Gestaltung**  
Technologieunternehmen



**Bei der Nutzung**  
Unternehmen, Behörden,  
Mitarbeitende, Privatpersonen etc.

**Bei der Entwicklung  
und Nutzung**  
Forschende,  
Forschungsinstitute etc.



- Aktive Auseinandersetzung
- Offener und kritischer Umgang
- Bewusstes Konsumverhalten
- Ausbau der Medienkompetenz

**Für Regulierung**  
Politische Akteure, Gesetzgeber etc.



# Anhang

## Papiere der Plattform Lernende Systeme zum Thema generative KI (eine Auswahl)

**Heesen, J.** (Hrsg.) (2020): Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen, Whitepaper aus der Plattform Lernende Systeme, München.

**Heesen, J. et al.** (2023): Künstliche Intelligenz im Journalismus. Potenziale und Herausforderungen für Medienschaffende, Whitepaper aus der Plattform Lernende Systeme, München.  
[https://doi.org/10.48669/pls\\_2023-1](https://doi.org/10.48669/pls_2023-1)

**Heesen, J., Müller-Quade, J., Wrobel, S. et al.** (Hrsg.) (2020): Zertifizierung von KI-Systemen – Kompass für die Entwicklung und Anwendung vertrauenswürdiger KI-Systeme, Whitepaper aus der Plattform Lernende Systeme, München.

**Heesen, J., Müller-Quade, J., Wrobel, S. et al.** (2020): Kritikalität von KI-Systemen in ihren jeweiligen Anwendungskontexten – Ein notwendiger, aber nicht hinreichender Baustein für Vertrauenswürdigkeit, Whitepaper aus der Plattform Lernende Systeme, München.

**Huchler, N.** (Hrsg.) (2020): Kriterien für die menschengerechte Gestaltung der Mensch-Maschine-Interaktion bei Lernenden Systemen, Whitepaper aus der Plattform Lernende Systeme, München.

**Löser, A., Tresp, V. et al.** (2023): Große Sprachmodelle – Grundlagen, Potenziale und Herausforderungen für die Forschung, Whitepaper aus der Plattform Lernende Systeme, München.

**Löser, A., Tresp, V. et al.** (2023): Große Sprachmodelle entwickeln und anwenden. Ansätze für ein souveränes Vorgehen, Whitepaper aus der Plattform Lernende Systeme, München.

**Plattform Lernende Systeme** (2019): Arbeit, Qualifizierung und Mensch-Maschine-Interaktion, Whitepaper der Arbeitsgruppe Arbeit/Qualifikation, Mensch-Maschine-Interaktion, München.

**Plattform Lernende Systeme** (2023): Hybride KI (Publikationsreihe KI Kompakt).  
[https://www.plattform-lernende-systeme.de/KI\\_Kompakt\\_Hybride\\_KI.pdf](https://www.plattform-lernende-systeme.de/KI_Kompakt_Hybride_KI.pdf)

**Plattform Lernende Systeme** (2024): AI Act der Europäischen Union. Regeln für vertrauenswürdige KI (Publikationsreihe KI Kompakt).  
[https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/KI\\_Kompakt](https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/KI_Kompakt)

**Samek, W., Schmid, U. et al.** (2025): Nachvollziehbare KI: Erklären, für wen, was und wofür. Plattform Lernende Systeme, München. DOI: [https://doi.org/10.48669/pls\\_2025-2](https://doi.org/10.48669/pls_2025-2)



# Webseite/Themenseite



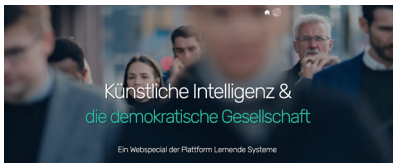
[GENERATIVE KI: CHATGPT & CO](#)



[DAS BRINGT GENERATIVE KI FÜR UNTERNEHMEN](#)



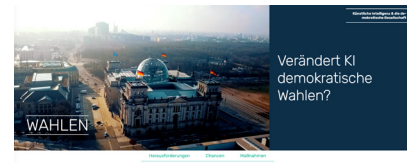
[AI ACT UND MEHR: REGELN FÜR KÜNSTLICHE INTELLIGENZ](#)



[KI UND DEMOKRATIE](#)



[KI UND JOURNALISMUS](#)



[KI UND WAHLEN](#)



[INTERVIEW PROF. DR. VOLKER TRESP](#)



[INTERVIEW ANNE LAUBER-RÖNSBERG](#)



[INTERVIEW ALEXANDER LÖSER](#)



[WEBTALKS & VERANSTALTUNGEN](#)

[NULL UND EINS MACHT KUNST? | MARC-UWE KLING | KONFERENZ 2025](#)

## Beratende Mitglieder der Plattform Lernende Systeme

**Eva Bittner**, Universität Hamburg

**Susanne Boll**, Carl von Ossietzky Universität Oldenburg

**Aljoscha Burchardt**, Deutsches Forschungsinstitut für Künstliche Intelligenz GmbH (DFKI)

**Björn Eskofier**, Friedrich-Alexander-Universität Erlangen-Nürnberg

**Axel Hahn**, Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut Systems Engineering für zukünftige Mobilität

**Jessica Heesen**, Eberhard Karls Universität Tübingen

**Andrea Martin**, IBM Watson Center Munich

**Rahild Neuburger**, Ludwig-Maximilians-Universität München

**Peter Rost**, Secunet Security Networks AG

**Ute Schmid**, Otto-Friedrich-Universität Bamberg

**Stefanie Schwaar**, Hochschule für Technik und Wirtschaft Berlin

**Andrea Stich**, Infineon Technologies AG

**Oliver Suchy**, Deutscher Gewerkschaftsbund DGB

## Redaktion

**Dr. Paul Grünke**, Geschäftsstelle der Plattform Lernende Systeme

**Alexander Mihatsch**, Geschäftsstelle der Plattform Lernende Systeme

**Christine Wirth**, Geschäftsstelle der Plattform Lernende Systeme

## Impressum

### Herausgeber

Lernende Systeme –  
Die Plattform für Künstliche Intelligenz  
Geschäftsstelle | c/o acatech  
Karolinenplatz 4 | 80333 München  
[www.plattform-lernende-systeme.de](http://www.plattform-lernende-systeme.de)

### Gestaltung und Produktion

PRpetuum GmbH, München

### Stand

November 2025

### Bildnachweis

Freepik / alle Illustrationen  
iStock / Trifonov\_Evgeniy, Gorodenkof / Seite 9

### Empfohlene Zitierweise

Plattform Lernende Systeme (2025):

**Zukunft gestalten! Mit generativer KI.**  
Gesellschaftliche Auswirkungen und  
Handlungsansätze

[https://doi.org/10.48669/pls\\_2025-7](https://doi.org/10.48669/pls_2025-7)

Bei Fragen oder Anmerkungen zu dieser Publikation kontaktieren Sie bitte Thomas Schmidt (Leiter der Geschäftsstelle): [info@plattform-lernende-systeme.de](mailto:info@plattform-lernende-systeme.de)

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, der Entnahme von Abbildungen, der Wiedergabe auf fotomechanischem oder ähnlichem Wege und der Speicherung in Datenverarbeitungsanlagen, bleiben – auch bei nur auszugsweiser Verwendung – vorbehalten.



