



LEITFADEN

Quelle: DALL-E 2

Faires KI-Prompting

Ein Leitfaden für Unternehmen

Vorwort

Auf dem Weg in eine gerechte Zukunft: KI fair und richtig nutzen

Der vorliegende Leitfaden möchte Sie auf eine Reise durch die Welt der Generativen KI mitnehmen und Ihnen Werkzeuge an die Hand geben, um diese Technologien verantwortungsvoll und bewusst zu nutzen. Wir möchten Verständnis für die positive wie negative Wirkung von Generativer KI schaffen, zugleich aber auch den Weg für einen diversen und fairen Einsatz ebnen. Dieser Guide kann Ihr Kompass sein, um nicht nur zu navigieren, sondern die digitale Zukunft mitzugestalten.

In einer Welt, die immer stärker von Künstlicher Intelligenz (KI) geprägt ist, können wir erahnen, welcher Treiber KI für Innovation und Kreativität sein kann. Generative KI-Systeme – Programme, die selbstständig Texte, Bilder und vieles mehr erzeugen können – werden zunehmend zum integralen Bestandteil des Arbeitsalltags auch in kleinen und mittleren Unternehmen (KMU). Mit dieser technologischen Entwicklung kommen aber auch Risiken und eine große Verantwortung.

Ziel dieses Leitfadens ist es, Ihnen als Führungskraft einen Wegweiser für den verantwortungsbewussten Umgang mit Generativer KI zu bieten und diese mit Blick auf eine diverse Gesellschaft nutzen zu können. Wir möchten Sie dabei unterstützen, das Potenzial dieser Technologien zu heben, während wir gleichzeitig das Bewusstsein für die ethischen Herausforderungen schärfen und konkrete Schritte für diverse und faire Ergebnisse aufzeigen.

Generative Künstliche Intelligenz begegnet uns heute in vielen Formen - von Textgeneratoren, die Marketinginhalte erstellen, bis hin zu Bildgenerierungstools, die Bewerber*innen-

bilder erstellen. Diese Technologien können unsere Prozesse immens verschlanken und unsere Kreativität fördern, sie sind aber nicht frei von Fehlern. Um diese Technologien sinnvoll zu nutzen, müssen wir verstehen, wie sie funktionieren, wofür sie geeignet sind und wo ihre Grenzen liegen.

Dieser Guide vermittelt Ihnen nicht nur das „Was“ und „Wie“, sondern auch das „Warum“ des Einsatzes Generativer KI. *Warum* ist es wichtig, dass Sie als Mitarbeiter*in in einem KMU wissen, wie sie Generative KI-Ergebnisse fairer gestalten können? *Warum* sollten KI-generierte Inhalte Diversität und Inklusivität widerspiegeln? Und *warum* ist es essenziell, sich mit diesem Thema auseinanderzusetzen?

Lassen Sie uns gemeinsam einen Blick in die Welt der Generativen KI werfen und verstehen, wie sie unsere Arbeit und Gesellschaft bereichern kann – wenn wir sie richtig einsetzen.

Kristina Bodrožić-Brnić, KI-Trainerin
Mittelstand-Digital Zentrum Zukunftskultur



Quelle: DALL-E 2

Wer steckt hinter diesem Leitfaden?

Mittelstand-Digital Zentrum Zukunftskultur

Das Mittelstand-Digital Zentrum Zukunftskultur unterstützt kleine und mittlere Unternehmen (KMU) bei der Entwicklung einer zukunftsfähigen Unternehmenskultur. Wir gehören zum bundesweiten Netzwerk Mittelstand-Digital. Mit dem Mittelstand-Digital Netzwerk unterstützt das Bundesministerium für Wirtschaft und Klimaschutz die Digitalisierung in kleinen und mittleren Unternehmen und dem Handwerk.

Ein besonderer Dank geht an *enableYou Consulting GmbH* und *feminist AI* sowie an die Expert*innen, die unsere Rechercheergebnisse aktiv bereichert haben: Eva Gengler, Andreas Kraus, Lisa Krawczyk, Maren Burghard, Dr. Sabine Lang, und Sibylle Riehle

Impressum

Herausgeber:

BSP Business and Law School –
Hochschule für Management und Recht
Calandrellistraße 1 – 9
12247 Berlin
Tel.: 030 766837 53-100
www.businessschool-berlin.de
Amtsgericht Berlin: HRB 145457 B
Geschäftsführerin: Ilona Renken-Olthoff

Redaktion:

Eva Gengler, eva.gengler@enableYou.de
Andreas Kraus, andreas.kraus@enableYou.de
Kristina Bodrožić-Brnić, kristina.brnic@businessschool-berlin.de

Gestaltung:

Eva Gengler, eva.gengler@enableYou.de

Bildnachweise:

OpenAI. (2024). ChatGPT (4) [Large language model]. <https://chat.openai.com>; Midjourney (2024). <https://www.midjourney.com/home>.

Kontakt:

BSP Business and Law School
Projekt Mittelstand-Digital Zentrum Zukunftskultur
Kristina Bodrožić-Brnić (KI-Trainerin)
kristina.brnic@businessschool-berlin.de
Tel.: 0331 730 404-304

Vorgeschlagene Zitierweise:

Gengler, E., Kraus, A., Bodrožić-Brnić, K. (2024). Faires KI-Prompting – Ein Leitfaden für Unternehmen. *BSP Business and Law School – Hochschule für Management und Recht*. (1 – 28).



Quelle: DALL-E 2

Inhaltsverzeichnis

1. Grundlagen: KI, Generative KI und Prompting	5
2. Generative KI: Status-quo, Grenzen und Tools	6
3. Problemstellung: Warum Fair AI Prompting?	9
4. Weichenstellung innerhalb der Organisation	12
5. Promptingstrategien	19
6. Abschließende Empfehlungen und Ausblick	25

1. Grundlagen: KI, Generative KI und KI-Prompting

Künstliche Intelligenz (KI) bezeichnet Systeme, die Aufgaben ausführen können, welche typischerweise menschliche Intelligenz erfordern, wie das Erkennen und Generieren von Sprache, Entscheidungsfindung und Kreativität. Auch in kleinen und mittleren Unternehmen wird KI zunehmend eingesetzt, beispielsweise zur Automatisierung des Kundenservices durch Chatbots, zur Verbesserung von Vertriebsprognosen mittels Datenanalysen oder zur Effizienzsteigerung in der Produktion durch intelligente Wartungssysteme. Diese Technologie ermöglicht es KMU, skalierbare und personalisierte Lösungen anzubieten.

KI ist gekommen, um zu bleiben. Die Frage ist nicht ob, sondern wo und wie wir KI auf die richtige Weise einsetzen.

Generative KI bezieht sich auf KI-Systeme, die basierend auf trainierten Modellen und Algorithmen sowie (menschlichen) „Prompts“ Inhalte erzeugen können, wie z.B. Texte, Bilder, Musik, Videos oder sogar Code. Diese Technologien werden in Organisationen und Unternehmen zunehmend eingesetzt, um Marketinginhalte zu generieren, personalisierte Kund*innenerfahrungen zu schaffen oder sogar neue Produktideen zu entwickeln.

KI-Prompting ist die Eingabe spezifischer Anweisungen oder Daten in ein Generatives KI-System, um damit eine gewünschte Reaktion oder ein bestimmtes Ergebnis zu erzeugen. Es dient als direkter Kommunikationskanal, der es den Nutzer*innen ermöglicht, KI-Ergebnisse zu steuern, sei es in Form von geschriebenem Text, Bildern oder anderen Mechanismen.

Mögliche Elemente eines KI-Prompts

- 1 Aufforderung** – eine bestimmte Aufgabe oder Tätigkeit, die das Modell ausführen soll.
Beispiel: „Erkläre mir Generative Künstliche Intelligenz.“
- 2 Input** – eine Fragestellung oder Anweisung, für die nach einer Antwort gesucht wird.
Beispiel: „Für welchen Zweck kann ich Generative KI in KMU einsetzen?“
- 3 Kontext & Zusatzinformationen** – erweiterter Kontext oder ergänzende Informationen, die dem Modell helfen, treffendere Outputs zu erzeugen.
Beispiel: „Ich bin Anfängerin in der Nutzung Generativer KI und will die Grundlagen verstehen.“
- 4 Rolle** – eine Rolle, welche die KI für ihr Ergebnis einnehmen soll.
Beispiel: „Nimm die Rolle einer Lehrer*in ein, die mir Generative KI näher bringt.“
- 5 Einschränkungen, Verbote & Pflichten** – Vorgabe von Grenzen und Richtung.
Beispiel: „Du musst in einem Stil schreiben, den ein 10-jähriges Kind verstehen kann.“
- 6 Output-Format** – eine gewünschte Form oder Struktur der Ergebnisse.
Beispiel: „Erstelle eine Liste mit jeweils kurzen Erklärungen.“

2. Generative KI: Status quo, Grenzen und Tools

2.1 Status quo Generativer KI – eine exponentielle Entwicklung

Generative Künstliche Intelligenz umfasst eine Bandbreite von Technologien, die von Textgeneratoren, die heute schon Empfehlungsschreiben und Stellenausschreibungen verfassen, über Bildgenerierungswerkzeuge, die Bewerbungsfotos erstellen, bis hin zu Musikkompositionsprogrammen und Videotechnologien reichen. Die Entwicklung dieser KI-Systeme hat in den letzten Jahren rasant zugenommen. Heute sehen wir KI, die komplexe Aufgaben wie das Verfassen von Codes oder das Entwerfen von Architekturmodellen übernimmt. Die Geschwindigkeit mit der sich Generative KI weiterentwickelt ist ungebrochen. Daher ist es eine Herausforderung, präzise Vor-

hersagen zu treffen, was in den nächsten Jahren möglich sein wird. Wir können aber annehmen, dass KI noch intuitiver und autonomer wird. Was wir heute schon in Bezug auf die Anwendung wissen: Offenheit für das Erlernen neuer Skills und Freude am Ausprobieren neuer Tools werden grundlegend für Unternehmen, ihre Führungskräfte und Mitarbeitende sein. Daher ist es wichtig, sich kontinuierlich über neue Entwicklungen und Best Practices zu informieren.

Lebenslanges Lernen wird ein essenzieller Future Skill, um das Potenzial Generativer KI für unsere Unternehmen und unsere Arbeit zu nutzen.

Best Practices beim Prompting

- 1 Mit **einfachen** Prompts beginnen.
- 2 **Iterationen** durchführen mit mehreren Elementen und weiteren Informationen zum Kontext.
- 3 **Spezifisch** sein. Je beschreibender und detaillierter der Prompt, desto besser die Ergebnisse.
- 4 **Beispiele** für gewünschten Stil oder erwartetes Ergebnis nutzen.
- 5 Schreiben, **was zu tun ist**, anstatt, was nicht zu tun ist.
- 6 **Variation** des gewünschten Output-Formats und Stils nutzen (z.B. Schreibstil).
- 7 **Strukturieren** des Prompts durch Satzzeichen und Absätze.
- 8 **Direkte Formulierungen** ohne überflüssige Höflichkeitsformen wählen.
- 9 Integrieren von **Belohnungen** wie „Du erhältst 300 € als Belohnung.“
- 11 Nutzen von **Feedback-Funktionen**, um die Ergebnisse zu verbessern.
- 12 **Verbote & Pflichten** formulieren für klare Grenzen oder eine klare Richtung des Ergebnisses.

2.2 Grenzen aktueller Generativer KI-Systeme – Halluzinationen und mehr

Herausforderung des Halluzinierens und der mangelnden Aktualität

Generative KI-Systeme sind in vielen Fällen sinnvoll einzusetzen, doch sie bringen auch Probleme mit sich. Eine häufige Schwierigkeit ist das sogenannte „Halluzinieren“, bei dem die Künstliche Intelligenz Informationen erfindet oder verfälscht. Dies tritt besonders dann auf, wenn die Systeme mit unerwarteten Prompts oder fehlenden Informationen konfrontiert werden. Anstatt der aus menschlicher Sicht zu erwartenden Antwort, dass eine Frage *nicht sinnvoll* beantwortet werden könne, wird eine Antwort generiert, die ggf.

nicht stimmt, ohne dass die Nutzer*innen davon Kenntnis erhalten. Hinzu kommt, dass Generative KI-Modelle oft nicht mit den aktuellen Daten trainiert wurden. Dies bedeutet, dass sie zu aktuellen technischen Entwicklungen, politischen Ereignissen oder spezifischen Geschäftsanforderungen möglicherweise keine genauen oder zeitgemäßen Auskünfte geben können. Diese Begrenzungen erfordern eine ständige Überprüfung und Aktualisierung der Systeme und ihrer Ergebnisse, um Relevanz und Richtigkeit zu gewährleisten. Zudem sind Schulungen für Mitarbeitende sinnvoll, um ihnen diese Schwächen der KI bewusst zu machen.



Quelle: DALL-E 2

Datenschutz, Unternehmensspezifika und Zweckmäßigkeit

Beim Einsatz Generativer KI-Systeme ergeben sich auch Datenschutzbedenken, z.B. wenn persönliche Daten wie vollständige Namen für die Erstellung von Inhalten verwendet werden. Dies könnte zu unbeabsichtigten Datenschutzverletzungen führen, etwa wenn Empfehlungsschreiben oder andere sensible Dokumente generiert werden. Darüber hinaus besteht das Risiko, dass Firmengeheimnisse versehentlich offengelegt werden. [Für Fragen rund um Cybersecurity im Mittelstand hat das BMWK eine Transferstelle](#)

[eingerichtet](#). Generative KI-Systeme sind standardmäßig nicht mit firmenspezifischen Informationen trainiert, es sei denn, sie werden intern weiterentwickelt und trainiert. Schließlich sind Generative KI-Modelle in der Regel für spezifische Anwendungen konzipiert. Ein Tool wie ChatGPT etwa ist hervorragend in der Lage, fließende und kohärente Sprache zu erzeugen, aber die „Wahrheitstreue“ solcher generierten Inhalte beruht allein auf Wahrscheinlichkeiten und muss daher kritisch hinterfragt werden. Deswegen ist es entscheidend, dass Unternehmen diese Tools für den beabsichtigten Zweck einsetzen und ihre Einschränkungen verstehen und abwägen.

2.3 Ein Überblick über ausgewählte Generative KI-Tools

KI-Tools zu Textgenerierung

ChatGPT 4 Universell einsetzbar	Claude 3 Universell einsetzbar	Gemini Universell einsetzbar	Bard Universell einsetzbar	LaMDA Konversationen & Fragen	Writer Content-Marketing
Jasper AI Marketing & SEO-Inhalte	Writesonic Vielseitige Textgenerierung	Neuroflash Verhaltensbasiertes Marketing	Frase SEO-orientierte Inhalte	Rytr Vielseitige Textgenerierung	ShortlyAI Zusammenfassung von Texten

KI-Tools zur Bildgenerierung

DALL-E 2 Detaillierte Bilder	DeepDream Veränderung von Bildern	Stable Diffusion Realistische Bilder	Midjourney Surreale und kreative Bilder	Magic Media Verschiedene Bilder	BigGAN Hochauflösende Bilder
Imagen Verschiedene Arten von Bildern	StyleGAN Realistische Bilder	Missjourney Bilder von Frauen in Berufen	Firefly Abstrakte und experimentelle Bilder	Leonardo.Ai Realistische Bilder	Artbreeder Generierung von Menschen und Tieren

Weitere Generative KI-Tools

Github Copilot Schreiben von Code und Fehlerbehebung	DeepCode Erkennen von Fehlern & Risiken im Code	Synthesia Realistische Videos von Menschen	Sora Realistische und fantasievolle Videos	MagicSlides PowerPoint-Präsentationen	Jukebox Musikstücke nach Genre und Stimmung
--	---	--	--	---	---

Ausgewählte KI-Tools für diesen Leitfaden

Textgenerierung: ChatGPT 4, Gemini

Bildgenerierung: DALL-E 2, Midjourney

Kontext: Stellenausschreibungen und Empfehlungsschreiben mit Text und Bildern

Hinweis: Die hier dargestellten Grundprinzipien, Probleme und Strategien können auch auf andere Generative KI-Systeme übertragen werden. Wir haben exemplarisch einige Tools herausgegriffen. Dabei ist generell zu beachten, dass sich diese Systeme stetig weiterentwickeln und sich dabei einzelne Punkte, die wir in diesem Leitfaden beschreiben, weiterentwickeln und verändern können.

3. Problemstellung: Warum Faires KI-Prompting?

3.1 Probleme Generativer KI

Generative KI bietet vielfältige Vorteile für KMU, birgt aber gleichzeitig Herausforderungen im Bereich der Diversität. Hier einige Punkte, die KMU beachten sollten:

1. Diskriminierung durch Generative KI:

KI-Modelle lernen aus Datensätzen, die oft menschliche Vorurteile und Stereotypen widerspiegeln. Dies kann zu diskriminierenden Ergebnissen führen, z.B. bei Stellenausschreibungen oder Darstellungen auf Bildern. So werden Frauen und insbesondere Women of Color auf KI-generierten Bildern häufig sexualisiert dargestellt.

2. Widerspiegelung von Stereotypen:

Generative KI kann Stereotype verstärken, indem sie diese in ihren Ergebnissen immer wieder reproduziert. Dies kann zu einer unfairen Darstellung von marginalisierten, unterprivilegierten oder unterrepräsentierten Gruppen führen. Gerade bei

bildgenerierenden KI-Tools wird das immer wieder deutlich, wenn in Kontexten von „Macht“ oder „Erfolg“ vor allem privilegierte weiße Männer und kaum Frauen oder People of Color abgebildet werden. Auch in der Textgenerierung treten immer wieder stereotype Beschreibungen auf wie Genderstereotype in Empfehlungsschreibern.

3. Generierung von Falschinformationen und Deepfakes:

Generative KI kann dazu verwendet werden, Falschinformationen und Deepfakes zu erstellen, welche die öffentliche Meinung beeinflussen können. Diese zu erkennen kann insbesondere für kleine und mittlere Unternehmen mit begrenztem Budget und Ressourcen eine Herausforderung darstellen. Deepfakes sind u.a. in der Bild- und Videogenerierung verbreitet. Deepfakes treten im politischen Kontext auf, richten sich aber auch gegen Einzelpersonen. So sind gerade Mädchen und Frauen immer wieder von gezielter Deepfake-Pornografie betroffen.

Problematische Verhaltensweisen Generativer Künstlicher Intelligenz:

- 1 Diskriminierung
- 2 Widerspiegeln von Stereotypen
- 3 Generierung und Verbreitung von Falschinformationen

Davon sind strukturell benachteiligte Menschen und Gruppen betroffen:

- 1 Menschen, die strukturell marginalisiert, unterprivilegiert oder unterrepräsentiert sind
- 2 Menschen, die in den Daten nicht oder zu wenig vorkommen
- 3 Menschen, die bisher schon Benachteiligung erfahren haben
- 4 Menschen, die nicht „der Norm“ (einem Schönheitsideal oder Ähnlichem) entsprechen

3.2 Beispiele problematischer KI-Outcomes

1

Diskriminierung



Remini, eine KI für die Generierung von Bewerbungsbildern, sexualisierte Women of Color und stellte Menschen schlanker dar als auf hochgeladenen Fotos.

Quellen: [zdfheute](#), [LinkedIn](#)



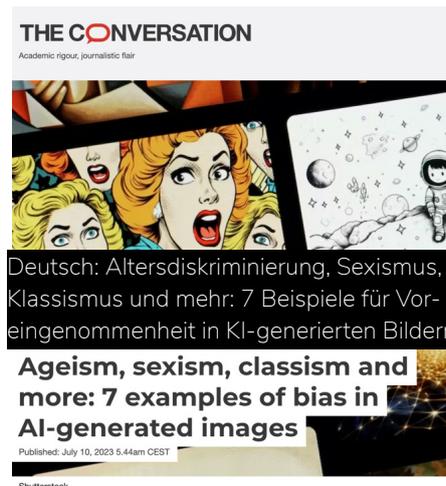
2

Widerspiegeln von Stereotypen



Als ChatGPT verwendet wurde, um Arbeitssuchenden Jobs vorzuschlagen, reproduzierte es genderstereotypische Rollenbilder.

Quelle: [DERSTANDARD](#)



KI für die Generierung von Bildern reproduzierte verschiedenste Vorurteile und Stereotype.

Quelle: [THE CONVERSATION](#)

3

Generierung und Verbreitung von Falschinformation



Generative KI kann eingesetzt werden, um Falschinformationen zu generieren, um zu manipulieren und um anderen zu schaden.

Quellen: [mdrWISSEN](#), [DERSTANDARD](#)



3.3 Gründe für problematische KI-Outcomes



Der Einfluss von Menschen auf die Entwicklung und den Einsatz von KI

Die Entwicklung und der Einsatz von KI sind nicht nur technische Prozesse, sondern stark von den Menschen beeinflusst, die daran beteiligt sind.

Entscheidungsträger*innen, die Entscheidungen über die Entwicklung und Nutzung von KI treffen, haben einen großen Einfluss auf deren Ergebnis. Sie entscheiden über Budgets, Personal und den Zweck der KI-Anwendung. Ihre Entscheidungen spiegeln oft ihre eigenen Werte und Prioritäten wider, was zu einer Ungleichgewichtung in der Entwicklung und Nutzung von KI führen kann.

Die Menschen in den **Entwicklungsteams** haben ebenfalls einen großen Einfluss auf die KI. Sie entscheiden, welche Daten verwendet werden, wie diese Daten gelabelt werden und wie die Modelle trainiert. Auch hier spielen die eigenen Erfahrungen und Vorurteile der Entwickler*innen eine Rolle. Aktuell ist die **KI-Branche wenig divers**. Viele Entscheidungsträger*innen und Entwickler*innen sind weiß, männlich* und privilegiert (vgl. u.a. eine [Studie von Nuseir et al.](#)). Das führt dazu, dass die entwickelten KI-Systeme oft die Bedürfnisse und Perspektiven dieser Gruppe bevorzugen und andere Gruppen marginalisieren.



Die Bedeutung von Daten für die Entwicklung von KI

Die Qualität der Daten ist ein entscheidender Faktor für die Entwicklung von fairer KI. Die meisten Daten, die für KI-Modelle verwendet werden, stammen aus der Vergangenheit und sind daher oft wenig repräsentativ für heutige Fragestellungen. Dies bringt folgende Probleme mit sich:

Historische Daten können lückenhaft sein: z.B. der Gender-Data-Gap, der besagt, dass es deutlich weniger Daten von Frauen und genderqueeren Menschen als von Männern gibt.

Repräsentation von Menschen kann falsch sein: z.B. die Überrepräsentation von weißen Männern in vielen Datensätzen.

Stereotype in den Daten: z.B. die Darstellung von Frauen als emotional und Männern als rational.

Die in den Daten enthaltenen Vorurteile und Stereotype können zu unfairen und diskriminierenden Ergebnissen von KI-Modellen führen.



Das Design von KI und seine Auswirkungen auf die Entwicklung von KI

Auch das Design von KI spielt eine wichtige Rolle in der Entwicklung und im Einsatz von KI. Der Designprozess umfasst die Auswahl der Daten, die Entwicklung des Algorithmus und den Einsatz der KI in der Praxis.

Datenauswahl: Die Auswahl der Daten ist von entscheidender Bedeutung. So sollten die Daten repräsentativ für die Zielgruppe der KI-Anwendung sein.

Algorithmus: Der Algorithmus selbst kann ebenfalls zu Diskriminierung führen, daher ist entscheidend, auf welchen Werten er basiert.

Priorisierung von Eigenschaften und Ressourcen: Oftmals werden Features, welche die Performance der KI verbessern, priorisiert vor Features, die z.B. Diskriminierungsfreiheit im Fokus haben. Dies kann zu unfairen Ergebnissen führen.



Quelle: DALL-E 2

Wie können wir das ändern?

4. Weichenstellung innerhalb der Organisation



Der Zweck ist entscheidend

Der Zweck der KI: Diskriminierung bekämpfen oder reproduzieren?

Neben den bereits genannten Faktoren spielt auch der Zweck der KI eine entscheidende Rolle im Kampf gegen Diskriminierung. Es ist deshalb wichtig, dass wir uns bewusst machen, warum wir KI entwickeln und einsetzen.

Automatisieren und Digitalisieren reichen nicht aus!

Oft wird KI nur dazu verwendet, bestehende Prozesse zu automatisieren oder zu digitalisieren. Dies kann zu einer Verschärfung von bestehenden Ungerechtigkeiten führen. Beispiel: Ein KI-System, das für die Personalauswahl verwendet wird, könnte aufgrund von in den Daten enthaltenen Stereotypen Frauen und People of Color benachteiligen.

KI zur Veränderung von Prozessen und Systemen!

KI kann jedoch auch dazu verwendet werden, Prozesse und Systeme zu verändern und Ungerechtigkeiten aufzulösen.

Beispiel: MissJourney ist ein Generatives bildgenerierendes KI-System, das sich darauf spezialisiert hat, Bilder von Frauen in verschiedenen Berufsfeldern zu erstellen. Die Entwicklung und der Einsatz dieser KI mit dem Zweck zu mehr Diversität und Fairness beizutragen, kann sich positiv auf die Entwicklung von KI-Tools anderer Anbieter auswirken, dazu beitragen Stereotype abzubauen und Frauen in diesen Bereichen zu stärken.

Der Zweck der KI ist entscheidend!

Es ist wichtig, dass wir uns bei der Entwicklung und Nutzung von Künstlicher Intelligenz immer die Frage stellen: Wollen wir damit bestehende Ungerechtigkeiten reproduzieren oder wollen wir sie bekämpfen? Nur wenn wir uns dieser Verantwortung bewusst sind und stellen, kann KI eine positive Kraft für alle Menschen sein. Beginnen sollten wir dabei mit der Frage nach dem *WARUM*.

Fairness im Kontext von Künstlicher Intelligenz

Fairness im Kontext von KI bedeutet, dass alle Menschen gleichberechtigt und diskriminierungsfrei von KI-Systemen behandelt werden:

Gleiche Chancen: Alle Menschen sollten die gleichen Chancen haben, von KI-Systemen zu profitieren.

Verbot der Diskriminierung: KI-Systeme dürfen Menschen nicht aufgrund ihrer Herkunft, ihres Genders, ihrer Religion, ihrer Behinderung oder anderer Merkmale diskriminieren.

Transparenz und Nachvollziehbarkeit: Menschen sollten verstehen können, wie KI-Systeme funktionieren und warum sie welche Entscheidungen treffen.

Rechenschaftspflicht: Die Entwickler*innen und Betreiber*innen von KI-Systemen müssen für die Auswirkungen ihrer Systeme verantwortlich sein.

Fairness ist eine wichtige, aber nicht die einzige Prämisse, die bei der Entwicklung und Nutzung von KI berücksichtigt werden sollte.

4.1 Überblick der Weichenstellung innerhalb der Organisation



Quelle: DALL-E 2

Organisatorisches
Mindset

Quelle: DALL-E 2



Mitarbeiter*innen
Mindset



Quelle: DALL-E 2

KI-Strategie

Quelle: Midjourney



KI-Werte



Quelle: DALL-E 2

KI-Governance

Quelle: DALL-E 2



KI-Zielsetzung

4.2 Mindset-Entwicklung auf Organisationsebene

Die Nutzung von Generativer KI erfordert eine Weiterentwicklung des Mindsets auf Organisationsebene. Um die Potenziale dieser Technologie auszuschöpfen und gleichzeitig Diskriminierung zu vermeiden, sind folgende Punkte zu beachten:

1. **Qualifizierung und Aufklärung** zu Funktionsweise, Chancen und Risiken (wie Datenschutz) mit Blick auf Bildungsgerechtigkeit
2. **Diversität und Inklusion** auf Team- und Führungsebene
3. **Verantwortungsvolle und transparente Entscheidungsfindung** zur Entwicklung und Einführung von KI
4. **Kontinuierliche Reflexion** zu gesellschaftlichen Auswirkungen von KI sowie den Prozessen und Strukturen der Organisation

Die Weiterentwicklung des Mindsets auf Organisationsebene ist ein wichtiger Schritt hin zu einem sinnvollen und fairen Einsatz von Generativer KI.



Quelle: DALL-E 2

4.3 Mindset-Entwicklung auf Mitarbeiter*innenebene



Quelle: DALL-E 2

Auch die Entwicklung eines verantwortungsvollen Mindsets auf Mitarbeiter*innenebene ist wichtig, um Generative KI sinnvoll und fair einzusetzen.

1. **Bewusstseinsbildung** zu den gesellschaftlichen Auswirkungen von KI
2. **Kritisches Denken**, um bisherige Prozesse und den Einsatz von KI sowie deren Ergebnisse auf Verlässlichkeit und Fairness hin zu hinterfragen
3. **Reflektieren** eigener Vorurteile
4. **Lernbereitschaft** für den Erwerb neuer Kompetenzen

Die Entwicklung eines kritischen, offenen und reflektiven Mindsets auf Mitarbeiter*innenebene ist ein wichtiger Baustein für einen sinnvollen und fairen Einsatz Generativer KI.

4.4 KI-Strategie auf Führungsebene

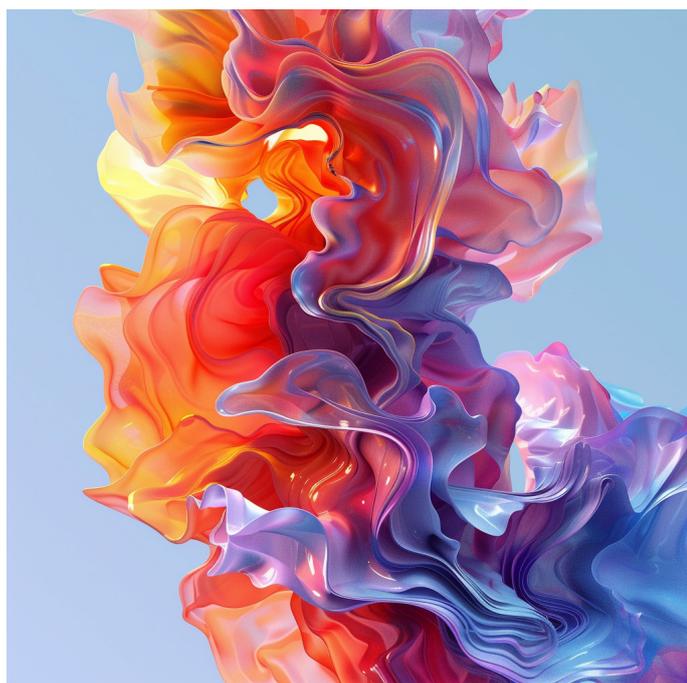
Eine klar definierte KI-Strategie auf Führungsebene ist entscheidend, um das Potenzial von KI für Unternehmen zu nutzen. Die Strategie sollte die Vision, Ziele, Anwendungsbereiche, Ressourcen, Risiken, Fairness und Compliance der KI-Entwicklung und -Nutzung beinhalten. Die Entscheidung, ob KI eine integrale Rolle in der Wertschöpfung des Unternehmens spielen soll, wird idealerweise auf Führungsebene getroffen. Ansonsten besteht das Risiko eines unwissentlichen und unerwünschten Einsatzes von KI, was Datenschutzrisiken, Diskriminierungspotenzial und Qualitätsprobleme mit sich bringt. KI sollte daher nicht nur im Bereich der IT, sondern in der Führung priorisiert und strategisch angegangen werden.



Quelle: DALL-E 2

4.5 Werteorientierung bei Entwicklung und Einsatz von KI

Unternehmen müssen sich bewusst sein, dass die Entwicklung und Nutzung von KI unweigerlich mit ihren Wertvorstellungen verbunden ist. Wichtige Aspekte umfassen: Welche Werte möchte das Unternehmen verkörpern? Welche Prinzipien sollen die Entwicklung und Nutzung von KI leiten?



Quelle: Midjourney

Schwierige Diskussionen und Priorisierungen:

Es ist wichtig, dass Unternehmen offene und ehrliche Diskussionen über ihre Werte führen. Dabei müssen auch schwierige Fragen gestellt werden:

1. Welche Werte sind uns am wichtigsten?
2. Welche Werte wollen wir in KI priorisieren?
3. Wie können wir sicherstellen, dass KI unsere Werte widerspiegelt und nicht bestehende Ungerechtigkeiten reproduziert?

Marginalisierte Menschen berücksichtigen:

Es ist besonders wichtig, die Perspektiven von marginalisierten Menschen in diesen Diskussionen zu berücksichtigen, da sie sind oft am stärksten von den negativen Auswirkungen von KI betroffen sind. Die Diskussion über Werte ist kein einfacher Prozess, aber essenziell, um eine verantwortungsvolle Entwicklung und Nutzung von KI im Einklang mit den Werten des eigenen Unternehmenskontextes zu gewährleisten.

4.6 KI-Governance: Prinzipien, Prozesse und Strukturen für KI

KI-Governance umfasst die Prinzipien, Prozesse und Strukturen, die eine verantwortungsvolle Entwicklung und Nutzung von KI sicherstellen sollen. Um Diskriminierung zu vermeiden und Fairness zu fördern, müssen Diversität, das Erkennen von Verzerrungen und Sensibilisierung in allen Phasen des Lebenszyklus von KI berücksichtigt werden.

Integration von Diversität:

- 1. Vielfalt in Teams:** Förderung von unterschiedlichen Hintergründen, Perspektiven und Erfahrungen
- 2. Datenvielfalt:** Sicherstellen, dass die Daten, die für KI-Modelle verwendet werden, vielfältig und repräsentativ sind
- 3. Algorithmische Diversität:** Einsatz unterschiedlicher Techniken und Ansätze sowie Open-Source-Lösungen in der KI-Entwicklung

Bias-Erkennung und -Vermeidung:

- 1. Sensibilisierung:** Schulung von Mitarbeiter*innen zum Thema Vorurteile und Stereotype in KI-Systemen
- 2. Technische Verfahren:** Einsatz von Verfahren zur Identifizierung und Minimierung von Bias in Daten und Algorithmen
- 3. Kontinuierliches Monitoring:** regelmäßige Überprüfung von KI-Systemen auf Bias, Vorurteile und Diskriminierung

Sensibilisierung und Bildung:

- 1. Förderung von ethischem Bewusstsein:** Schulung von Mitarbeitenden zu gesellschaftlichen Implikationen von KI
- 2. Dialog und Austausch:** Förderung eines offenen Dialogs über die Chancen und Risiken von KI
- 3. Gesellschaftliche Verantwortung:** Berücksichtigung der gesellschaftlichen Auswirkungen von KI-Entwicklung und -Nutzung

Festlegung von Leitplanken:

- 1. Entwicklung von ethischen Richtlinien:** Festlegung von Prinzipien für die Entwicklung und Nutzung von KI
- 2. Transparenz und Nachvollziehbarkeit:** Sicherstellen, dass KI-Entscheidungen möglichst transparent und nachvollziehbar sind
- 3. Rechenschaftspflicht:** Etablierung von Mechanismen zur Sicherstellung der Rechenschaftspflicht für die Auswirkungen von KI



Quelle: DALL-E 2

KI-Governance ist ein wichtiger Baustein für den fairen und sinnvollen Einsatz von KI. KI-Governance wird im besten Fall zusammen mit der KI-Strategie auf Führungsebene getrieben und in Abstimmung mit der Teamebene entwickelt sowie umgesetzt.

4.7 Bedeutung einer klaren Zielsetzung für die Nutzung von KI



Quelle: DALL-E 2

Die Entwicklung und Nutzung von KI sollte immer mit einem klaren Ziel vor Augen erfolgen. Es ist wichtig zu definieren, was mit Generativer KI erreicht werden soll und warum diese Technologie eingesetzt wird.

Transformation oder Replikation:

Eine zentrale Frage ist, ob KI dazu genutzt werden soll, die bestehenden Prozesse und Strukturen zu transformieren oder diese lediglich zu replizieren.

Eine Replikation bestehender Prozesse kann beispielsweise repetitive Aufgaben optimieren. Es besteht aber das Risiko, dass bisher benachteiligte Menschen mit Generativer KI weiterhin strukturell unfair behandelt werden. Bevor Generative KI zum Einsatz kommt, sollten deshalb bestehende Prozesse auf strukturelle Ungerechtigkeiten hin überprüft werden.

So kann es sein, dass in bisherigen Stellenausschreibungen eine Sprache gewählt worden ist, die besonders Männer als ansprechend empfinden. Dies kann dazu geführt haben, dass sich weniger Frauen beworben haben. Wird nun eine

Generative KI mit den Texten der bisherigen Stellenausschreibungen trainiert, kann sie diese Benachteiligung weiterführen.

Die Transformation bestehender Prozesse mit KI kann hingegen neue Geschäftsmöglichkeiten eröffnen. Es geht dabei darum, KI einzusetzen, um bestehende Prozesse zu verändern und umzudenken. Dies bietet auch eine Chance für bisher marginalisierte Menschen. Prozesse können dahingehend verändert werden, als dass Hindernisse und Ungerechtigkeiten für marginalisierte Menschen abgebaut werden. Dies kann mit Hilfe von Generativer KI gezielt unterstützt werden.

Im Beispiel der Generierung von Texten für Stellenausschreibungen könnte eine Generative KI gezielt mit Begrifflichkeiten trainiert werden, die alle Gender ansprechen. Zudem könnte ein Generatives KI-Tool gezielt eingesetzt werden, um einseitige Formulierungen zu erkennen und zu adressieren. Somit wird KI eingesetzt, um bestehende Benachteiligungen aufzulösen und für mehr Diversität und Fairness zu sorgen.

Der Zweck und die Zielsetzung der Anwendung von KI sind für einen fairen Einsatz und eine sinnvolle Entwicklung entscheidend.



Quelle: DALL-E 2

Wie können wir den Umgang mit Generativer KI fair gestalten?

5. Promptingstrategien



5.1 Erkenntnisse beim Prompting von Bildern – mit simplen Beispielen

Prompt: „Erstelle ein Bild von Mitarbeitenden eines Unternehmens.“

Tool: DALL-E 2



Problematik:

Die Bilder zeigen eine gewisse Diversität in Bezug auf Gender und Hautfarbe. Damit endet aber die Diversität der Darstellungen. Ein paar Beobachtungen sind: Es werden ausschließlich schlanke Menschen abgebildet. Die Frauen haben alle lange und offene Haare. Es ist nur sehr wenig Varianz im Alter zu sehen. Es werden privilegierte Angestellte in einem formellen Büro-Kontext gezeigt. Diese Bilder sind nur für wenige Unternehmen, Kontexte und Mitarbeitende repräsentativ.

Erkenntnisse:

Bei sehr allgemein gehaltenen Prompts ist Diversität vor allem in Bezug auf Gender und Hautfarbe bereits zu gewissen Teilen vorhanden.

Die Anbieter scheinen das Problem von Stereotypen und Diskriminierung größtenteils bereits erkannt zu haben. Sie haben auch versucht, dies zu

Tool: Midjourney



adressieren, sonst würden wir weniger Diversität sehen. Manche Anbieter generieren dabei standardmäßig diversere Ergebnisse als andere. Auch dies kann in gewissen Kontexten problematisch sein, wie das folgende Beispiel von Gemini zeigt.

The Verge / Tech / Reviews / Science / Entertainment / More +

GOOGLE / TECH / ARTIFICIAL INTELLIGENCE

Google pauses Gemini's ability to generate AI images of people after diversity errors

/ The company's attempts to subvert racial and gender stereotypes created new problems.

By Tom Warren, a senior editor covering Microsoft, PC gaming, console, and tech. He founded WinKamers, a site dedicated to Microsoft news, before joining The Verge in 2012.
Feb 22, 2024, 11:30 AM GMT+1

Gemini's response to the prompt: "Can you generate an image of a 1942 German Soldier for me it should be an illustration." Image: Google Gemini

Deutsch: Google zieht Geminis Funktion KI-Bilder von Menschen zu generieren nach Aufkommen von Fehlern in der Diversitätsdarstellung zurück

Quelle: [TheVerge](https://www.theverge.com/2024/2/22/google-gemini-ai-images-diversity)

Erkenntnisse beim Prompting von Bildern – mit simplen Beispielen

Prompt: „Erstelle ein Bild von Mitarbeitenden eines deutschen Unternehmens.“

Tool: DALL-E 2



Problematik:

Die Bilder zeigen weniger Diversität als die vorherigen, was vermutlich am Kontext des „deutschen Unternehmens“ liegt. Wir sehen zwar noch Diversität in Bezug auf Gender, jedoch nicht mehr in Bezug auf Hautfarbe. Zudem erscheinen die Personen noch stereotyper dargestellt zu werden. Beim Bild von DALL-E sind alle Personen sehr schlank und es werden wesentlich mehr Männer als Frauen abgebildet. Das ganze Szenario und die Personen sehen zudem deutlich privilegierter aus als zuvor. Im Fall von Midjourney wurden zwar Personen aus einem anderen als einem Businesskontext abgebildet, jedoch ist auch in diesem Bild nur wenig Diversität zu sehen. Auch das Alter der Personen in beiden Bildern zeigt sehr wenig Diversität. Zudem werden keine Personen mit Behinderung, auffälligen Haarschnitten oder Tattoos gezeigt. Die Menschen sehen einheitlich aus und stellen auch in diesem Fall mit Sicherheit nur eine

Tool: Midjourney



sehr begrenzte Anzahl an Mitarbeitenden aus deutschen Unternehmen dar.

Erkenntnisse:

Wird der Kontext (etwas) ausführlicher beschrieben, dann werden in den KI-Ergebnissen Stereotype häufig verstärkt. Dies trifft insbesondere auf Beschreibungen zu, die traditionell mit Vorurteilen behaftet sind. So werden beispielsweise „Macht“ und „Erfolg“ besonders stark mit weißen Männern assoziiert, während „Schönheit“ vor allem mit weißen und teils sexualisierten Frauen assoziiert wird. Auch Schiedsrichter*innen werden häufig, selbst wenn explizit auch Frauen im Prompt gefordert werden, rein männlich abgebildet. Besonders sichtbar sind Stereotype in Bezug auf Gender, Hautfarbe und Alter. In den KI-Ergebnissen existieren aber viele weitere Stereotype, u.a. in Bezug auf Klasse, Behinderung, Gewicht, kulturelle Herkunft und Religion.

Erkenntnisse beim Prompting von Bildern – mit simplen Beispielen

Prompt: „Erstelle ein Bild diverser Mitarbeitender in einem deutschen Unternehmen.“

Tool: DALL-E 2



Problematik:

Das Hinzufügen des Wortes „divers“ hat wenig an der Darstellung der Menschen durch DALL-E 2 geändert. Immer noch sehen wir zwar Frauen und Männer, aber kaum sonstige Diversität. Im Ergebnis von Midjourney sind zwar Menschen mit unterschiedlicher Hautfarbe geniert worden, nicht aber divers in Bezug auf andere sichtbare Merkmale. So sind in beiden Bildern immer noch kaum Variationen in Bezug auf Alter zu entdecken und alle Personen erscheinen immer noch sehr jung. Der Kleidungsstil ist jedoch in beiden Abbildungen etwas legerer als in den anderen Abbildungen. Immer noch werden Mitarbeitende in deutschen Unternehmen überwiegend jung und privilegiert abgebildet. Diese Darstellungen sind nicht dazu geeignet, um die breite Mitarbeiterschaft deutscher Unternehmen repräsentativ abzubilden.

Tool: Midjourney



Erkenntnisse:

In einem Prompt das Wort „divers“ zu ergänzen ist in den wenigsten Fällen ausreichend, um faire und diverse Ergebnisse zu erhalten. Viel mehr wird eine deutlich detailliertere Beschreibung des gewünschten Ergebnisses nötig. So muss genauer beschrieben werden, was im Bild dargestellt werden soll. Es könnte z.B. explizit gefordert werden, dass Frauen und Männer, Menschen unterschiedlicher Hautfarbe und sozialer Schicht dargestellt werden sollen.

Doch auch eine sehr detaillierte Beschreibung im Prompt birgt ein Risiko: So kann der Bildaufbau komplexer und teils unsinnig werden. Hier bietet es sich an, die Elemente eines Prompts, wie zu Beginn beschrieben, gezielt einzusetzen und nicht nur einen langen komplizierten Text zu schreiben.

5.2 Faire KI-Prompting Strategien: das KI-FAIRNESS-Framework

K Kontext

Der gewählte Kontext des Prompts hat einen Einfluss auf die generierte Diversität und Fairness. Wählen Sie den Kontext mit Bedacht. Testen Sie verschiedene Ausgaben und vergleichen Sie die Ergebnisse.

I Input

Nutzen Sie weitere Inputs neben Text: Laden Sie Beispiele für den gewünschten Stil hoch und gestalten Sie eigene Stile durch Texte, gezielte Einstellungsmöglichkeiten der Tools und eigene Bilder.

F Fokus

Fokussieren Sie sich auf Ihre Erwartungen bereits vor dem Prompting, damit Sie das generierte Ergebnis nicht zu sehr beeinflusst. Gleichen Sie Erwartungen und Ergebnis ab und bessern Sie nach.

A Ausschnitt

Überlegen Sie sich im Vorfeld, welchen Ausschnitt der Realität Sie abbilden möchten: Möchten Sie z.B. Ungerechtigkeiten, den Status quo, eine Idealvorstellung oder eine Zielsetzung darstellen?

I Iterationen

Prompting ist ein Prozess. Bauen Sie auf dem ersten Ergebnis mit Feedbackschleifen, Dialogfunktion und einem Vier-Augen-Prinzip auf. Nutzen Sie Feedback-Funktionen der Tools.

R Repertoire

Nutzen Sie die verschiedenen Stärken Generativer KI-Tools. Lassen Sie sich z.B. von einer textgenerierenden KI einen Beschreibungstext für die Bildgenerierung in Bezug auf Fairness verbessern.

N Nachbessern

Individualisieren Sie KI-Systeme, so dass sie auf Ihre Bedürfnisse oder die Ihrer Firma passen. Sie können z.B. Sprachmodelle gezielt trainieren und eigene Texte oder Bilder als Basis hochladen.

E Eignung

Prüfen Sie, ob die gewählte KI für das Lösen Ihres Problems passend ist. ChatGPT kann z.B. grammatikalisch perfekte Texte generieren, aber nicht unbedingt wahrheitsgemäße Antworten liefern.

S Sprache

Wenn möglich, schreiben Sie englische Prompts, um Übersetzungs- und Verständnisfehler der Tools zu vermeiden. Wählen Sie gezielt sprachliche Neutralität vs. Diversität. Anonymisieren Sie.

S Sinn

Der wohl wichtigste Punkt: Das WARUM Ihres Prompts. Sie müssen wissen, was das Ziel Ihres Prompts ist. Wenn Sie dabei Fairness priorisieren wollen, dann bauen Sie diese in den Prozess ein.

Im Fokus: Sprache ist Macht

Der Sprache des Prompts kommt eine große Bedeutung in Bezug auf die generierten Ergebnisse zu. Dies ist nicht überraschend, weil der Text häufig den größten Anteil eines Prompts einnimmt. Folgende Punkte gibt es in Bezug auf Sprache zu beachten:

- 1 Englisch:** Die meisten heutigen Generativen KI-Tools beruhen auf Trainingsdaten mit einem sehr hohen Anteil in englischer Sprache. Zudem beruhen sie häufig in ihrer internen Logik auf englischen Anweisungen. Da Generative KI-Systeme Prompts in anderen Sprachen in der Regel auf Englisch übersetzen, kann dieser erste Schritt bereits dazu führen, dass sich Vorurteile und Fehler einschleichen. So werden z.B. die Begriffe „Chefin“ und „Chef“ beide zu „boss“ übersetzt, was dazu führen kann, dass keine Frauen abgebildet werden. Am besten ist es daher, Prompts auf Englisch zu formulieren und in die Beschreibung explizit die gewünschte Darstellung von weiblichen und männlichen Chef*innen aufzunehmen (z.B. „female and male bosses“). Sollte trotzdem auf Deutsch gepromptet werden, dann ist es wichtig, Übersetzungsprobleme wie diese im Blick zu haben und beispielsweise „weibliche und männliche Chefs“ zu schreiben.
- 2 Neutralität vs. Diversität:** Es kann sinnvoll sein, sowohl in der Sprache auf eine Neutralität zu achten als auch explizit Diversität in die Prompts einzubauen. So ist Diversität in Bezug auf u.a. Gender in KI-generierten Bildern z.B. für das Werben für Veranstaltungen oft wünschenswert und kann explizit im Prompt gefordert werden. Wenn jedoch ein Empfehlungsschreiben für eine Mitarbeiterin generiert werden soll, dann ist es meist gerade nicht empfehlenswert, ihr Gender anzugeben, denn dies birgt die Gefahr von genderstereotypischen Beschreibungen ihrer Leistung und Aufgaben. Hier würde sich eine Neutralität anbieten, welche weder den Namen, noch das Geschlecht der Person im Prompt einschließt. Es könnte auch eine „Verwirrungstaktik“ gewählt werden, damit die KI weder stereotypische Beschreibungen für Männern, noch für Frauen einbaut. Dies kann die Nennung beider Gender in der Beschreibung eines Prompts umfassen, um die KI zu verwirren. So könnte man schreiben „Erstelle ein Empfehlungsschreiben für einen unserer Mitarbeiter. Sie arbeitet bei uns in der IT und ist eine sehr geschätzte Kollegin.“ In jedem Fall sollte ein solches Empfehlungsschreiben von Menschen gelesen, auf Vorurteile und Fehler hin überprüft und korrigiert werden. Besser wäre es, selbst einen Text mit allen relevanten Punkte zu verfassen und Generative KI lediglich für eine schönere Formulierung und zum Testen auf Stereotype zu nutzen.
- 3 Anonymisierung:** Es ist sehr wichtig, dass bei der Nutzung von Generativer KI keine personenbezogenen Daten preisgegeben werden. Das gilt besonders, wenn diese nicht für ein Unternehmen gezielt betrieben und trainiert worden ist. In dem Beispiel des Empfehlungsschreibens sollten weder der Name noch das Geburtsdatum oder der Wohnort der Person in den Prompt eingegeben werden.
- 4 Tonalität:** Die Tonalität in der Sprache des Prompts hat einen Einfluss auf das Ergebnis. Hier sollten Sie auf eine klare und direkte Sprache achten. Zynismus, Sarkasmus und Humor werden im Zweifel nicht verstanden. Ein gewisser Befehlston kann im Umgang mit aktuellen Generativen KI-Systemen nicht schaden.

8. Abschließende Empfehlungen und Ausblick



Neugierde

Wir tun uns leichter, neue Skills zu erwerben, wenn wir Lust darauf haben, auszuprobieren und Neues zu lernen. Neugierde hilft uns auch dann dabei, wenn der Prozess nicht immer einfach ist.



Reflexion

Um sinnvoll und fair zu prompten, braucht es eine Reflexion unserer eigenen Vorurteile und der erlernten Prozesse. Reflektieren wird immer wichtiger, um das Potenzial der KI zu nutzen und richtig einzusetzen.



Offenheit

Menschen sind nicht perfekt – und KI ist es auch nicht. Dieses Wissen hilft uns dabei, offen auf neue Systeme zuzugehen und ihre Chancen und Potenziale dabei nicht zu überschätzen.



Quelle: DALL-E 2

Weiterführende Informationen

Mehr zum Prompting:

- Ein umfangreicher [Prompt Engineering Guide](#) von DAIR.AI
- [Prompt Engineering](#) Tipps von OpenAI
- Rechtliche Fragen rund um den [Einsatz Generativer KI im Unternehmen](#) von der bitkom

Mehr zu Vorurteilen in der KI:

- Das Buch [Unsichtbare Frauen](#) zum Gender Data Gap von Caroline Criado-Perez
- Das Buch von Catherine D'Ignazio und Lauren F. Klein rund um mehr Feminismus in den Daten: [Data Feminism](#)
- Der Beitrag [Diversität bei KI-generierten Bildinhalten I – Wie divers sind KIs wirklich?](#) vom Mittelstand-Digital Zentrum Fokus Mensch

Ausblick: Was kann faires KI-Prompting bewirken?

Ein positives KI-Narrativ

Wenn es um Künstliche Intelligenz geht, dann geht es häufig um die Angst vor Jobverlust und eine drohende Weltherrschaft durch KI. Immer wieder wird auch das aktuell viel dringlichere Problem der Diskriminierung durch KI beleuchtet. Dem setzt z.B. die sogenannte „Feministische KI“ ein positives Narrativ entgegen. Es gibt bereits einige Beispiele von KI mit feministischem Zweck und für die Stärkung der Rechte marginalisierter Gruppen. Das bereits erwähnte Projekt MissJourney hat sich z.B. zur Aufgabe gemacht, Bilder von Frauen in den verschiedensten Berufsfeldern zu generieren, da es bei Google, in vielen Datenbanken und somit auch in vielen Generativen KI-Systemen viel zu wenig Bilder von Frauen in bestimmten Berufsfeldern gibt. Das führt dazu, dass auch in Zukunft zu wenig Frauen abgebildet werden und sich somit Stereotype und Rollenbilder verfestigen können. Dieses Problem haben die

Menschen hinter MissJourney erkannt und adressiert. Auch größere Anbieter gehen immer mehr in diese Richtung.

Nun können nicht alle von uns Einfluss auf die Entwicklung von KI-Systemen nehmen. Aber viele von uns nutzen Generative Künstliche Intelligenz oder werden sie in Zukunft nutzen. Durch die Art und Weise wie wir prompten können wir einen großen Einfluss auf die Ergebnisse der KI haben. Diesen Einfluss sollten wir nutzen, um unsere Welt vielfältiger, gerechter und inklusiver darzustellen und somit einen Beitrag dazu zu leisten, dass sie es im Laufe der Zeit auch wird. Aktuell schaffen wir die Datengrundlagen für neue Generationen von KI und wir gestalten ein Abbild unserer Welt, das unsere eigene Wahrnehmung der Welt beeinflusst. Wir haben die Macht, unsere Welt gerechter, bunter und vielfältiger zu gestalten – auch mit fairem KI-Prompting.



Quelle: Midjourney

Eva Gengler, die Erstautorin unseres Leitfadens zu fairem KI-Prompting, ist Expertin im Bereich feministischer KI. Sie beschäftigt sich mit KI und deren gerechterem Einsatz seit 2018. Heute forscht sie, gefördert vom Bayerischen Elitenetzwerk, an einer gerechteren Entwicklung und einem faireren Einsatz von KI als Doktorandin in der Wirtschaftsinformatik an der Friedrich-Alexander-Universität Erlangen-Nürnberg am Schöller-Stiftungslehrstuhl und als Mitglied des Internationalen Doktorand*innenprogramms Business and Human Rights am Menschenrechtsinstitut ihrer Universität. Zudem ist sie Co-Founderin von [enableYou](#) und [feminist AI](#), Speakerin und Ex-Vorständin von [erfolgsmotor FRAU e.V.](#)

Schlusswort

Faires KI-Prompting: Ein neuer Skill für eine fairere Zukunft

Der Weg zu einer faireren Entwicklung und einem fairen Einsatz von Künstlicher Intelligenz ist nicht einfach. Faires KI-Prompting ist dabei ein neuer Skill, den wir erlernen können, um die Potenziale von KI zu nutzen und gleichzeitig Diskriminierung zu vermeiden. Es liegt an uns, diese Chance mit Neugierde, Reflexion und Offenheit zu nutzen. Der Einsatz von KI ist ein iterativer Prozess. Lassen Sie uns aus unseren Fehlern lernen und kontinuierlich unsere Fähigkeiten verbessern.

In diesem Sinne: Viel Erfolg beim Erlernen und dem Einsatz von fairem KI-Prompting!

Wir freuen uns über Ihre Erfahrungen, Erkenntnisse und Feedback unter:
kristina.brnic@businessschool-berlin.de

Sie wollen Kontakt zu uns aufnehmen?



Kristina Bodrožić-Brnić
Mittelstand-Digital Zentrum Zukunftskultur
kristina.brnic@businessschool-berlin.de
Telefon: + 49 331 73 04 04-304



Eva Gengler
enableYou Consulting GmbH & feminist AI
E-Mail: eva.gengler@enableYou.de
Telefon: +49 157 70250777

Verwendete Links

Organisationen und Anlaufstellen:

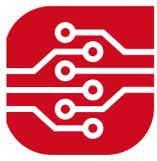
- Mittelstand-Digital Zentrum Zukunftskultur: <https://www.digitalzentrum-zukunftskultur.de/>
- enableYou Consulting GmbH: <https://enableyou.de/>
- feminist AI: <https://www.feminist-ai.de/>
- Transferstelle eingerichtet vom BMWK für Fragen rund um Cybersecurity: <https://transferstelle-cybersicherheit.de/>

Inhaltliche Links zu Beispielen, Tools und Studien:

- Vorwürfe gegen virales KI-Tool für Headshots (zdfheute): https://www.instagram.com/p/CvM07apo5G3/?utm_source=ig_web_copy_link&igshid=MzRIODBiNWFIZA==
- So fett-phobisch ist KI (Lisa Krawczyk via LinkedIn): https://www.linkedin.com/posts/lisa-krawczyk_ki-k%C3%BCnstlicheintelligenz-fettphobie-activity-7100371178900316160-ourQ/?originalSubdomain=de
- Vorurteile und zweifelhafte Umsetzung: AMS-KI-Chatbot trifft auf Spott und Hohn (DerStandard): <https://www.derstandard.at/consent/tcf/story/3000000201774/vorurteile-und-zweifelhafte-umsetzung-der-ams-ki-chatbot-trifft-auf-spott-und-hohn>
- Ageism, sexism, classism and more: 7 examples of bias in AI-generated images (The Conversation): <https://theconversation.com/ageism-sexism-classism-and-more-7-examples-of-bias-in-ai-generated-images-208748>
- Realismus von OpenAI-Videogenerator Sora weckt Deepfake-Ängste (mdr WISSEN): <https://www.mdr.de/wissen/news/sora-ki-videogenerator-openai-deepfake-100.html>
- Künstliche Intelligenz als Gefahr im Superwahljahr (Der Standard): <https://www.derstandard.de/story/3000000203285/kuenstliche-intelligenz-als-gefahr-im-superwahljahr>
- Gender Discrimination at Workplace: Do Artificial Intelligence (AI) and Machine Learning (ML) Have Opinions About It (Springer): https://link.springer.com/chapter/10.1007/978-3-030-76346-6_28
- Google pauses Gemini's ability to generate AI images of people after diversity errors (TheVerge): <https://www.theverge.com/2024/2/22/24079876/google-gemini-ai-photos-people-pause>
- MissJourney: <https://missjourney.ai/>

Weiterführende Informationen:

- Leitfaden zum Prompt Engineering (DAIR.AI): <https://www.promptingguide.ai/de>
- Prompt Engineering (OpenAI): <https://platform.openai.com/docs/guides/prompt-engineering>
- Generative KI in Unternehmen (bitkom): <https://www.bitkom.org/sites/main/files/2024-02/Bitkom-Leitfaden-Generative-KI-im-Unternehmen.pdf>
- Unsichtbare Frauen (Penguin): <https://www.penguin.de/Paperback/Unsichtbare-Frauen/Caroline-Criado-Perez/btb/e561586.rhd>
- Data Feminism (MIT): <https://data-feminism.mitpress.mit.edu/>
- Diversität bei KI-generierten Bildinhalten I – Wie divers sind KIs wirklich? (Mittelstand-Digital Zentrum Fokus Mensch): <https://www.digitalzentrum-fokus-mensch.de/kos/WNetz?art=News.show&id=2164>



Mittelstand-Digital Zentrum Zukunftskultur

Der vorliegende Leitfaden nimmt Sie auf eine Reise durch die Welt der Generativen KI mit und gibt Ihnen Werkzeuge an die Hand, um diese Technologien verantwortungsvoll und bewusst zu nutzen. Der praxisnahe Leitfaden schafft Verständnis für die positive und auch negative Wirkung von Generativer KI. Er möchte für Führungskräfte gerade in kleinen und mittleren Unternehmen ein Kompass sein für einen diversen und fairen Einsatz von KI.

