



LEITFADEN

Green Prompting Guide

Effizient arbeiten, Ressourcen schonen und präzise Prompten im Betriebsalltag

Gefördert durch:



Mittelstand-Digital 

aufgrund eines Beschlusses
des Deutschen Bundestages

Vorwort

Anwendungen aus dem Bereich der Generativen KI überfluten aktuell den Markt und werden in nahezu jede aktuelle Soft- & Hardware integriert. Menschen generieren mit den Modellen Texte, Audio, Grafiken, Bilder, Musik oder Videos.

Dieser Leitfaden unterstützt dabei, die neuen Anwendungen auf eine verantwortungsvolle Art und Weise zu nutzen: Green Prompting.

Dieses noch sehr junge Feld bietet schon heute vielfältige Möglichkeiten, nachhaltige Digitalisierung in den Alltag zu integrieren.

Viel Freude beim Durchstöbern wünscht Ihnen nun das Team des Mittelstand-Digital Zentrums Zukunftskultur und des Zukunftszentrum Berlin.

St

Impressum

Herausgeber:

Hochschule für nachhaltige Entwicklung Eberswalde
Schicklerstr. 5
16225 Eberswalde
zukunftskultur@hnee.de

Rechtsform:

Die HNEE ist eine Körperschaft des öffentlichen Rechts. Sie wird nach außen durch den amtierenden Präsidenten, Prof. Dr. Matthias Barth, vertreten. Die HNEE untersteht der Rechtsaufsicht des Ministeriums für Wissenschaft, Forschung und Kultur des Landes Brandenburg.

Bildnachweise:

Titelbild: eigene Darstellung;

Sonstige Bilder: Diverse Bildquellen, die jeweils unter dem jeweiligen Bild genannt werden.

Redaktion:

Susann Feuerschütz
Gerrit Neuhaus
Mittelstand-Digital Zentrum Zukunftskultur

Francesca Hermani
Zukunftszentrum Berlin

Gestaltung und Produktion:

Gerrit Neuhaus, Susann Feuerschütz & Francesca Hermani

Kontakt:

Susann Feuerschütz
susann.feuerschuetz@hnee.de
www.digitalzentrum-zukunftskultur.de

<https://www.digitalzentrum-zukunftskultur.de>

<https://www.zukunftszentrum-berlin.de>

Stand: Februar 2026

Green Prompting

Inhalt

| | |
|---|-----------|
| Der Green Prompting Guide | 4 |
| Der ökologische Fußabdruck der Generativen KI | 4 |
| Datengrundlagen | 7 |
| Warum ist das wichtig? | 9 |
| Guidelines für Green Prompting | 10 |
| Basics: Suffiziente Nutzung und Promptplanung | 11 |
| Maßvoller KI Einsatz | 13 |
| Prompt Planung | 17 |
| Upgrades: Generative KI noch grüner einsetzen | 24 |
| Nachhaltige Infrastruktur | 25 |
| Organisation & Zusammenarbeit | 27 |
| Ausblick | 30 |
| Literatur | 32 |

Indikatoren Generative

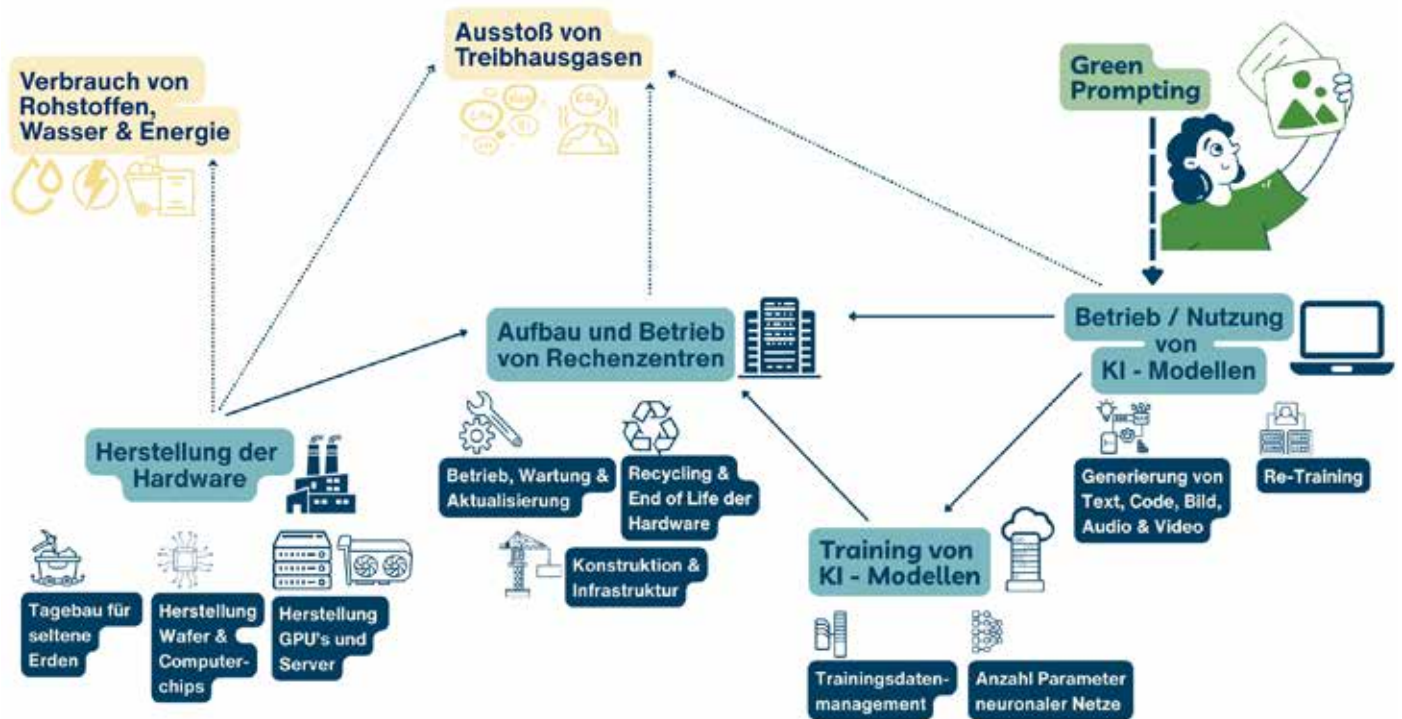
Was hat Generative KI eigentlich mit der Umwelt zu tun? Generative KI ist eine Form der Künstlichen Intelligenz, die auf Basis von Trainingsdaten selbstständig neue Inhalte wie Texte, Bilder, Videos oder Programmcodes erstellt.

Ökologie die GenAI meets

Auf den folgenden Seiten stellen wir den Zusammenhang von Umweltauswirkungen mit der Nutzung von Anwendungen Generativer KI als Grundlage für diese Guidelines in den Mittelpunkt.

Die Herstellung der Hardware für Chips, Endgeräte, Server, sowie Rechenzentren selbst verbraucht wertvolle Rohstoffe (Seltene Erden, Metalle, Kunststoffe, Sand, Öl etc.). Zumindest seltene Erden werden größtenteils in Konfliktregionen ohne jegliche Umweltstandards gewonnen. Auf die Herstellung von Chips und Hardware entfallen laut einer Studie der internationalen Energieagentur etwa 20% des Energieverbrauchs im Lebenszyklus der Generativen KI.

Der ökologische Lebenszyklus der Generativen KI



© Eigene Darstellung HNEE, 2026

Der Energieverbrauch, der während der Trainingsphase eines GenAI Modells anfällt, variiert in Abhängigkeit der Größe des Modells und der Komplexität der Hardwarekonfiguration des Rechenzentrums.

Die internationale Energieagentur hat in einer Versuchsanleitung den durchschnittlichen Energieverbrauch des Trainings eines großen KI-Modells mit etwa 1,7TWh berechnet.

Stromverbrauch des Trainings eines großen LLMs

Energieverbrauch des Trainings eines großen KI Modells **Jährlicher Stromverbrauch von etwa 1 Million Personen in Deutschland**

Nimmt man an, dass eine Person in Deutschland jährlich etwa 1.600 kWh Strom verbraucht, ergibt sich der Vergleich mit dem Stromverbrauch von etwa 1 Million Personen pro Jahr.



© Eigene Darstellung auf Basis von IEA 2025; HNEE, 2026

In der Nutzungsphase lösen bereits heute Milliarden täglicher Anfragen Rechenoperationen auf der Hardware von Rechenzentren aus. Ein Chat in dem Texte, Code, Bilder, Audio oder Video generiert werden, verbraucht deutlich mehr Strom als eine einfache Anfrage an eine Suchmaschine im Webbrowser. Es kann festgehalten werden, dass Anfragen mehr Energie und Ressourcen verbrauchen, je komplexer die generierten Inhalte sind.

Datengrundlagen

Wie groß der Einfluss unserer Prompts auf den ökologischen Fußabdruck ist, können wir nicht konkret sagen, solange wir Systeme von Drittanbietern nutzen. Entsprechende Zahlen werden nicht oder nur zurückhaltend - meist an Fachpublikum adressiert - veröffentlicht. Gleichzeitig schreitet die technische Entwicklung und somit die Effizienz von KI-Systemen, Computerchips und Rechenzentren so schnell voran, dass Informationen von heute, in wenigen Monaten schon nicht mehr aktuell sein können.

Wie viel Strom verbraucht eigentlich ein Sprachmodell? Diese Frage stellte die Nachhaltigkeitsmanagerin des Lebensmittelunternehmens Bauck Mühle ihrem IT-Verantwortlichen Michael Host. Im Austausch mit dem Mittelstand-Digital Zentrum Zukunftskultur diskutierte er dies und machte sich kundig. In einem Versuchsaufbau mit verschiedenen Sprachmodellen, Wattmeter als Stromzähler, Token-Zähler und GPU-Tracking zur Bewertung der Rechenleistung testete er Prompts.



© HNEE 2025, Versuchsaufbau Michael Host

Erkenntnisse aus dem Test:

- ✓ Kurze, klare Prompts = flache Leistungskurve
- ✓ Große Modelle = großer Durst
- ✓ Bewusste Sprache = messbar weniger Energieverbrauch



© HNEE 2025, Michael Host auf dem Barcamp - KI für eine bessere Welt ?!

Green Prompting klingt nach Theorie – bis man den Stromzähler danebenlegt. Und plötzlich wurde sichtbar, was sonst verborgen bleibt: Ein einziger Prompt kann deinen Stromzähler hochtreiben. Ein klar formulierter Prompt lässt ihn fast ruhig stehen.

Michael Host

<https://www.linkedin.com/feed/update/urn:li:activity:7391136770819710976/>

Dr. Friederike Rohde formuliert vor diesem Hintergrund eine Forderung an Politik und KI-Industrie. Wer sich den ganzen Vortrag zu Thema ansehen möchte, kann den QR Code oder folgenden Link nutzen.

Dr. Friederike Rohde vom Bits & Bäume Netzwerk bzw. dem Berlin Ethics Lab der TU Berlin skizzierte die vielschichtige Problematik des Energie- und Ressourcenverbrauchs von KI. Dieser verursache sowohl ökologische Folgen, z.B. durch immensen, ansteigenden Wasser- und Stromverbrauch, als auch soziale Konflikte z.B. um Landnutzung oder Verteilungsgerechtigkeit dieser Ressourcen. Denn sowohl der Bau von Rechenzentren als auch die Chipindustrie hätten, in Übersee und auch in Europa nicht nur ein großes „Ressourcenpreisschild“, sondern auch ein Transparenzproblem.

Wir müssen ein Gleichgewicht bei der Wissensverteilung über die Umweltfolgen Generativer KI herstellen. Die Informationen dürfen nicht in Konzernschubladen verschwinden!

Dr. Friederike Rohde



© HNEE 2025, Dr. Friederike Rohde auf dem Barcamp - KI für eine bessere Welt ?!

Hier den ganzen Videovortrag auf der Zusammenfassungsseite des „Barcamp - KI für eine bessere Welt?!“ ansehen.

<https://www.digitalzentrum-zukunftskultur.de/artikel/ki-verantwortungsvoll-einsetzen-14382/>



Der Nutzung von Drittanbieter Apps steht der Betrieb von KI-Systemen auf eigener Hardware entgegen, die eine transparentere Datengrundlage ermöglicht. So wurden im Rahmen einer Studie der Hochschule München aus dem Jahr 2024 in diversen Versuchsanordnungen unterschiedliche KI-Modelle untersucht, die auf der eigenen IT-Infrastruktur installiert wurden. Hier konnte anhand der Generierung verschiedener Inhalte der konkrete CO₂-Footprint ermittelt und dargestellt werden. Das Bild einer Katze mit Kappe und Brille vor dem Computer hat im Zuge der Studie bei seiner Erstellung 4,7 g CO₂-Äquivalente verursacht.



Dieses KI-generierte Bild verursachte bei der Entstehung 4,7 g CO₂-Äquivalente. Zwei Textfragen verursachen bei großen Sprachmodellen ähnlich viel. Quelle: Hochschule München (2025): Neue Studie zeigt: Je besser KI antwortet, desto größer ihr ökologischer Fußabdruck (Pressemitteilung)

Warum ist das wichtig für uns?

Wir wissen nicht genau, wie groß der Fußabdruck eines einzelnen Prompts ist. Es kann aber festgehalten werden, dass der Einfluss der Nutzungsphase auf den Fußabdruck der Generativen KI mit ansteigender Nutzung auch stetig zunimmt. Als Privatperson und in noch größerem Maße als Mitarbeitende in Unternehmen treffen wir viele Entscheidungen rund um KI: Wir entscheiden, ob oder wie wir die Anwendungen nutzen und treffen möglicherweise Kaufentscheidungen (z.B. Gen AI-Anbieter/Modell, Endgerät (Smartphone, Tablet, PC, Produktionsmaschine), Stromanbieter). Diese Entscheidungen wirken direkt auf die Umweltkosten. Wir verursachen sie also aktiv. Dies ist auch der Bereich, in dem wir Verantwortung übernehmen und unsere Umweltwirkungen optimieren können.

Die Entscheidungen, die wir beim Prompten von Gen AI treffen, liegen komplett in unserer Hand. Die eigene Handlungsmacht, zu entscheiden, wann und wie man Generative KI einsetzt, wollen wir in diesem Leitfaden in den Fokus stellen. Hierzu haben wir Guidelines in zwei Ebenen unterteilt. Im folgenden Abschnitt finden Sie die Basics mit denen Sie ihr Promptingverhalten bei der Nutzung der Systeme von Drittanbieter ausrichten können. Darauf folgen Upgrades, die Sie bei der Nutzung und Entwicklung eigener Systeme beachten können

Good News!

Schon kleine Änderungen leisten einen Beitrag. Positives Umwelthandeln und der Sinn für eine umsichtige, nachhaltige Nutzung von Technologie sind wertvolle Bestandteile einer modernen Unternehmenskultur und zukunftsfähigen Arbeitswelt.

Definiertes Guiding

Auf den folgenden Seiten geben wir einen Einblick in die Möglichkeiten durch das eigene Nutzungsverhalten Einfluss auf den ökologischen Fussabdruck der Generativen KI zu nehmen. Dabei gehen wir im ersten Teil den Basics auf grundlegende Promptingkonzepte ein. Unter den Upgrades erläutern wir weiterführende Strategien für eine umweltverträglichere Nutzung Generativer KI.

Basics & Upgrades

Basics: Suffiziente Nutzung und effiziente Planung

Wir starten mit ein paar grundlegenden Überlegungen zur Funktionsweise herkömmlicher Gen AI Modelle. Gen AI funktioniert wie ein Chatbot, ein computerbasiertes System, mit dem die Nutzenden in natürlicher Sprache interagieren. Die Sprachmodelle, die hinter der Chatbot-Oberfläche liegen, werden auf Basis riesiger Datenmengen trainiert, Muster zu erkennen. Die Ausgaben der Systeme sind keine Antworten im menschlichen Sinne, sondern wahrscheinlichkeitsbasierte Zufallsprodukte. Somit sind KI-Systeme keine Wahrheitsmaschinen. Es ist belegt, dass man eher mit einer systembedingten „Faktenuntreue“ rechnen muss: z.B. werden weitverbreitete Irrtümer als Antworten ausgegeben (sog. „Selection Bias“*). Zudem haben Antworten von Gen AI Chatbots grundsätzlich einen menschlichen Bias, qua Design und qua Trainingsdatenlage, und derzeit sogar durch gezielt ins Netz gestellte Falschinformationen. Menschliches Feedback ist eine Strategie, mit der GenAI Anbieter versuchen, dieses Problem abzumildern, gelöst ist es jedoch nicht.

*Bias bezeichnet kognitive Wahrnehmungsverzerrungen (z.B. Vorurteile, Stereotypen und andere Denkfehler)

Philipp Slusallek, wiss. Direktor des Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) ordnete KI folgendermaßen ein:

„Künstliche Intelligenz ist nicht neutral. Die Werte, Normen, Informationen der Menschen, die sie gebaut haben, stecken in dem System drin.“



© HNEE 2025, Prof. Dr. Philipp Slusallek auf dem Barcamp - KI für eine bessere Welt ?!

Der EU AI Act* sei ein großer Meilenstein im verantwortungsvollen Umgang mit KI. Trusted AI **würde Garantien liefern, was KI-Systeme können sollen, können dürfen und was nicht. Fehlendes Weltwissen von Gen AI ist ein „Makel“, der zu größeren, energieintensiveren Modellen führen werde. Slusallek rechne künftig mit destillierten Modellen, die auf Basis dieser „weiseren“ Riesensysteme auf Einzelaufgaben spezialisiert und verlässlicher würden.

*Der EU AI Act (kurz KI-Verordnung) ist das weltweit erste umfassende Gesetzespaket der EU zur Regulierung von Künstlicher Intelligenz. Es trat im August 2024 in Kraft und wird schrittweise bis 2026 verbindlich. Weitere Informationen finden Sie auf Seite 28.

** Trusted AI bezeichnet KI-Systeme, die transparent ethischen Grundsätzen folgen, gesetzliche Vorgaben erfüllen und technisch zuverlässig sind.

Hier den ganzen Videovortrag auf der Zusammenfassungsseite des „Barcamp - KI für eine bessere Welt?!“ ansehen.

<https://www.digitalzentrum-zukunftskultur.de/artikel/ki-verantwortungsvoll-einsetzen-14382/>



fiz

suf

Unter dem Ansatz Digitaler Suffizienz versteht Mensch den Verzicht auf den Gebrauch digitaler Medien ohne dabei auf den gewünschten Effekt verzichten zu müssen.

fiz

ienz

Maßvoller Einsatz Generativer KI

Diese Grenzen und Risiken der Generativen KI, v.a. Bias, mangelndes Weltwissen bis zu Halluzinationen, sollten vor allem in der professionellen Nutzung Berücksichtigung finden. Die wichtigste Grundlage für „Green Prompting“ ist daher nur dann zu Prompten, wenn ein klarer Zweck bzw. ein Ziel definiert ist und feststeht, dass dieses Ziel auch erreicht werden kann; es geht sozusagen um „Prompten mit Sinn und Verstand“.

Suffizienz

Die grundlegende Betrachtung eines maßvollen Einsatzes von Generativer KI leitet sich aus der Suffizienzstrategie der Nachhaltigkeit ab. Es geht um die bewusste Nutzung und das Weglassen unnötiger Prompts, um die entsprechende ökologische Belastung gar nicht erst entstehen zu lassen.

Tipp 1: Kein Prompt ist der „grünste“ Prompt

Wofür möchte ich die Gen AI einsetzen? Kann ich mein Ziel möglicherweise anders und sogar besser erreichen? Erhalte ich durch den Einsatz realistisch betrachtet ein schnelleres und / oder besseres Ergebnis?

Wenn ein Prompt nicht nötig ist, vermeiden Sie ihn besser. Das ist nicht nur gut für die Umwelt, sondern auch für die eigene Denkleistung. Ob mit oder ohne KI: Aufgaben kritisch zu reflektieren, ist kostenloses Training Ihrer Synapsen.

Nicht zu Prompten ist eine so einfache Maßnahme, dass sie auf den ersten Blick trivial wirkt. Doch: Genau zu wissen, wann prompten einen Mehrwert hat und wann es eigentlich Verschwendung ist, ist nicht ganz so leicht. Orientierung liefert vielleicht das Use-Case-Canvas des Mittelstand-Digital Zentrums Zukunftskultur.



Das Use-Case-Canvas finden Sie unter folgendem Link oder Sie folgen einfach dem QR-Code.

<https://www.digitalzentrum-zukunftskultur.de/material/ki-use-case-canvas-14860/>

Prüfung der bestehenden Rahmenbedingungen im Unternehmen

Oft hilft bereits ein Blick auf die Rahmenbedingungen, die beispielsweise in den eigenen Unternehmensrichtlinien, die hinsichtlich der KI Nutzung im Unternehmen oder zu den Themen IT Security und Datenschutz zur Verfügung stehen und erste Orientierung zur Nutzung liefern können. Zu den Themen Unternehmensrichtlinien gibt es eine Vielzahl von Angeboten der Mittelstand-Digital Zentren.



Den Beitrag zu Leitplanken für den Arbeitsalltag finden Sie unter folgendem Link oder Sie folgen dem QR-Code.

<https://www.digitalzentrum-zukunftskultur.de/material/ki-richtlinie-leitplanken-fur-den-arbeitsalltag-14754/>

Zielbezug

Gibt es keine Orientierung für Ihren Anwendungsfall, ist es hilfreich sich seines konkreten Anwendungsbereichs bewusst zu sein und ihr Ziel auf Basis ihrer beruflichen Herausforderungen im Blick zu behalten. Der Wunsch nach Steigerung von Produktivität und Effizienz kann auch zu einer Übernutzung Generativer KI führen ohne einen konkreten Nutzen zu ziehen.

Tipp 2 : Ziel im Auge behalten

Viele Gen AI Anwendungen machen nach einer Anfrage konkrete Angebote und Vorschläge, was sie als nächstes tun können.

Halten Sie kurz inne: Brauchen Sie diese weitere Interaktion wirklich oder ist Ihr Ergebnis gut genug, um selbst damit weiterzuarbeiten?

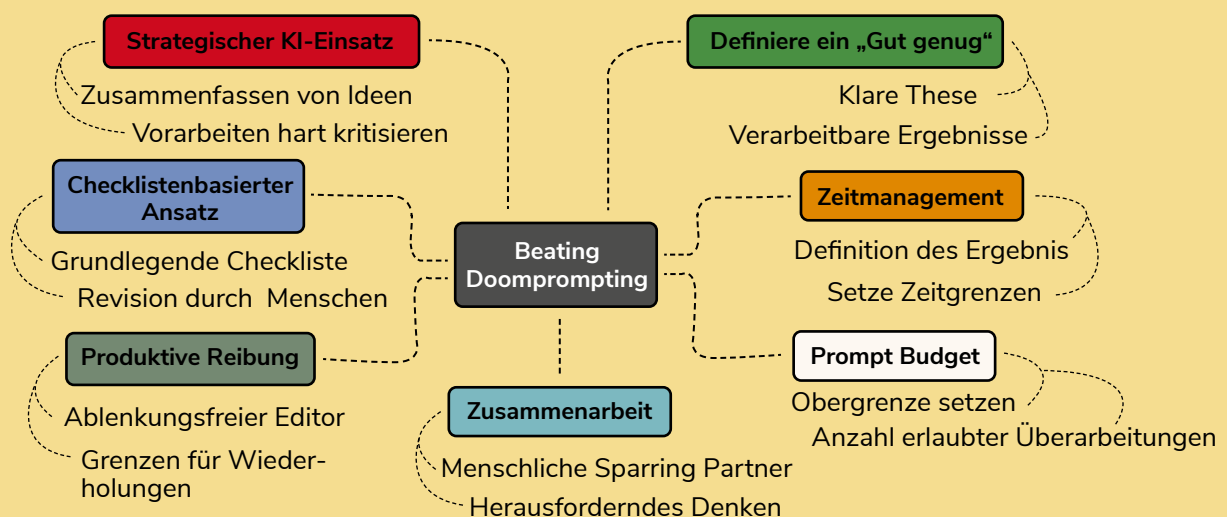
Setzen Sie sich ein inneres Limit: „Mehr als drei Prompts verwende ich für dieses Problem nicht. Dann suche ich einen anderen Weg.“

Exkurs aus der Fachwelt - Doomprompting

Als **Doomprompting** bezeichnet man das Phänomen, des „**Sich Verlierens im Prompten**“. Viele, wengleich nicht jede Gen AI bieten zusammen mit ihrer Antwort auf Prompts **weitere Vorschläge** an, was die KI nun mit den Ergebnissen tun könnte. Auch ohne diese Vorschläge gibt es Nutzende, die dazu neigen erste Ergebnisse **ständig zu verbessern oder anzupassen**, ohne jemals ihr gewünschtes Ziel / Ergebnis zu erzielen. Dies kann zu **Endlosschleifen und langen Interaktionen** mit vielen Prompts führen, die von vorneherein nicht geplant waren.

Ähnlich wie das Phänomen „**Doomscrolling**“, bei dem Nutzende sozialer Medien **incentiviert** werden, in dem Medium zu bleiben und **weiterzumachen**, verschwendet Doomprompting viele Ressourcen. Denn oft ist der **Mehrwert**, der entsteht, in dem man ein gutes Ergebnis noch ein wenig besser macht, **nicht mit dem Energieaufwand** oder der eigenen **Zeit** und **Produktivität** zu rechtfertigen. Die nachfolgende Grafik fasst Strategien gegen das Doom Prompting zusammen. Wer das Thema vertiefen möchte, folgt einfach dem Link unter der Grafik.

Strategien für maßvollen KI Gebrauch



© frei übersetzt nach: ABV Creative im Medium Magazin, 2024 [LINK](#)

Testumgebung

Natürlich müssen alle Menschen Erfahrungen mit der Technologie machen, ihre Grenzen kennenlernen, Kompetenzen erwerben und zudem über die richtigen Daten für die jeweilige Aufgabe verfügen. „KI Readiness“, fasst dies in einem Begriff zusammen, nämlich die Fähigkeit einer Organisation und ihrer Mitarbeitenden, mit KI produktiv und verantwortungsbewusst umzugehen. Im professionellen Kontext müssen daher auch Grundlagen für möglichen Gen AI Einsatz vorliegen (z.B. eine firmeninterne KI-Richtlinie, Schulungen) .

Unsere Empfehlung ist, zu Beginn des Einsatzes das Ausprobieren für Mitarbeitende zu fördern – im Rahmen von verantwortungsvollem Einsatz. Austausch über sinnvolle und wenig zielführende Prompts im Team sind förderlich. Doch auch die Sensibilisierung für die „Umweltkosten“ des Promptens sollte im Betrieb Platz haben.

Effi

Unter Digitaler Effizienz verstehen wir die zielgerichtete Erreichung oder Herstellung eines Nutzens mit möglichst geringem Ressourceneinsatz - das gilt z.B. für Geld, Zeit, Energie.

zie

nz

Prompt Planung

Sehr viel Energie können Sie sparen, indem Sie Prompts gut planen. Dieses sog. „Prompt Engineering“ wurde bislang vor allem mit Blick auf Genauigkeit der Ergebnisse diskutiert. Ein gut strukturierter Prompt kann jedoch messbar Energie sparen. Wenn die grundsätzliche Entscheidung für den sinnvollen Gen AI Einsatz getroffen wurde, geht es nun an die konkrete Umsetzung:

Erste Studien zeigen, dass die Antwortlänge ein wichtiger Faktor für den Energieverbrauch einer Gen AI – Anfrage ist.

Tipp 3 : Präzise und konkret Prompten

Überlegen Sie vor dem Prompten: Welches Ergebnis benötige ich?

Formulieren Sie die Aufgabe im Prompt semantisch präzise. Dies spart Energie durch weniger Rechenleistung und kürzere, fokussierte Antworten.

Vage Prompts wie: „Erzähl mir etwas über Nachhaltigkeit in der Automobilindustrie“ liefern lange, wenig brauchbare Antworten.

Besser: „Nenne Maßnahmen, mit denen ein KMU in der Automobilindustrie seine Lieferkette nachhaltiger gestalten kann.“

Die Länge des Prompts selbst spielt eine eher geringe Rolle, wichtiger ist, was der Prompt als Befehl für Wörter bzw. Inhalte enthält. Zudem gibt es erste Hinweise darauf, dass komplexere Aufgaben und bestimmte Stichworte zu längeren Antworten und höherem Energieaufwand führen:

Prompts mit den Befehlen „analysiere“ und „erkläre“ verbrauchten bspw. mehr Energie als Prompts mit „klassifiziere“ – dies wurde bei verschiedenen Modellen gemessen. Ähnlich verhält es sich auch mit Höflichkeitsfloskeln, die in der Summe ihrer Verwendung zu einem gesteigerten Energieverbrauch in der Nutzungsphase führen können.

„Lachs im Fluss“ - ein Quatsch-Ergebnis

Auch in der **Bildgenerierung** kann viel schiefgehen. Hier wurde, ein Foto zu dem **unpräzisen Prompt** „Lachs im Fluss“ erstellt.

Das ist zwar witzig, aber – ganz gleich, wie hoch der Energieverbrauch von Bildgenerierung ist, hier ist der Fall klar:

Ein **unbrauchbares Ergebnis** ist **Verschwendung** von Zeit, Geld & Energie.



© JazJon, Reddit 2022, [LINK](#)

Prompt Engineering

Im Bereich der Prompt Planung gibt es bestimmte Prompt Strategien, auch bekannt als „Prompt Engineering“. Doch welche Auswirkungen haben Techniken wie Zero-Shot-Prompting (Prompten ohne Beispiele), One- oder Few-Shot-Prompting (Prompten mit einem oder mehreren Beispielen) auf den Energieverbrauch?

Belastbare Ergebnisse gibt es hierzu bislang nicht, jedoch wächst das wissenschaftliche Interesse an diesen Fragen. Den größten Einfluss hat bei Tests in vorläufigen Studien die Größe des Gen-AI-Modells (vgl. Tipp 7). Lediglich für zwei Prompt-Engineering-Techniken gibt es Informationen zum Energieverbrauch: das energiesparende Batch-Prompting und das energieintensive Chain-of-Thought-Prompting.

Batch Prompting: Green Prompting Content Suite

System Persona & Context

„Du bist ein Experte für nachhaltiges Marketing und Green Tech. Dein Ziel ist es, Content zu erstellen, der maximalen Mehrwert bietet, während die Token-Ausgabe präzise und ohne unnötige Füllwörter bleibt (CO2-effizientes Prompting).“

Task Instruction

„Verarbeite die folgenden drei Content-Module basierend auf dem Thema „Green Prompting Guide“ in einem einzigen Durchlauf. Trenne die Ergebnisse klar durch horizontale Linien.“

Variable Input: [Thema: Green Prompting Guide]

Modul 1: LinkedIn-Post

Fokus: „Warum effiziente Prompts Rechenpower sparen.“

Stil: „Professionell, provokant, lösungsorientiert.“

Struktur: „Hook (Problem), Core (3 Tipps), Call-to-Action.“

Modul 2: Newsletter-Teaser

Fokus: „Ankündigung des neuen „Green Prompting Guide“.“

Stil: „Einladend, informativ.“

Vorgabe: „Maximal 150 Wörter.“

Modul 3: Instagram Captions

Variante A: „Quick-Tip (Kurz & Knapp).“

Variante B: „Myth-Buster (KI ist immer ein Stromfresser).“

Constraint-Regeln für maximale Effizienz:

Keine Einleitungstexte wie: „Hier sind deine Ergebnisse“.

Starte direkt mit Output.

Verwende Markdown für die Formatierung.

Token-Limitierung Halte die Antworten: „Dicht; Qualität vor Wortanzahl.“

Batch Prompting

Batch-Prompting eine Technik, bei der mehrere Abfragen in einer einzigen Prompteingabe zusammengefasst werden, sodass das Sprachmodell die Antworten für alle Abfragen in einem einzigen Durchgang generieren kann. Dieser Ansatz reduziert die Anzahl der API-Aufrufe (API Aufrufe sind das Zusammenspiel aus Anfrage, Bearbeitung sowie Antwort durch ein Large Language Model. API Aufrufe werden z.B. durch Prompteingaben ausgelöst.) und den gesamten Rechenaufwand.

Damit ist es eine effizientere Lösung für groß angelegte Anwendungen. Die Einsatzbereiche reichen von Kundensupport, Erstellen von Marketing-Content, Produktabfragen im Einkauf bis hin zur Datenanalyse.

Chain-of-Thought (CoT) Prompting

Chain-of-Thought (CoT) ist eine Technik, bei der das Modell durch eine Reihe von Argumentationsschritten geführt wird. Indem es eine komplizierte Aufgabe in kleinere Schritte oder „Argumentationsketten“ aufteilt, kann das Modell Sprache besser verstehen und genauere Ergebnisse liefern.

Grundsätzlich benötigt CoT mehr Energie und sollte daher sparsam, gezielt eingesetzt werden. Für komplexe Use Cases ist es empfehlenswert, wenn die Trefferquote beim ersten Prompt hoch ist und Sie Wiederholungen vermeiden.

Beispiel für einen Chain-of-thought Prompt

Sie wollen Indikatoren aus einem 20-seitigen Nachhaltigkeitsbericht ziehen (Daten aus unstrukturierten Texten). Bei der Few-Shot-Technik bräuchten Sie sehr viele Beispiele, ggf. auch mehrere Anläufe.

Mit CoT geben Sie diesen Prompt ein:

„Identifiziere erst alle Zahlen, filtere dann nach Umwelt-KPIs und berechne die Differenz zum Vorjahr.“



Einen Selbstlernkurs zu Grundlagen des Prompt Engineering finden Sie hier:

<https://lea.ita-kl.de/url/RepositoryEntry/76808196>

Tokens Sparen Begrenzungen einstellen

Ein weiterer Tipp ist es, Tokens zu sparen – dies ist, auf der Eingabeseite (Prompt) und der Ausgabe-seite (Ergebnis) möglich.

Ein **Token** ist in der Informatik die Einheit, die Algorithmen einer Gen AI compu-terisieren.

Anders ausgedrückt: KI verarbeitet Daten in der eigenen Sprache, die auf einem Vokabular von Tokens basiert. Tokens sind **Dateneinheiten**, die entstehen, wenn **größere Informationsblöcke** zerlegt werden.

Zum Vergleich: Zum Beispiel entspricht in GPT-Modellen 1 Token ungefähr 4 Zeichen oder 0,75 Wörtern im englischen Text.

Die **Geschwindigkeit und Energieeffizienz**, mit der Tokens verarbeitet werden, ist eine wichtige **Leistungsgröße** bei der Bewertung von KI-Modellen. Viele An-bieter arbeiten daran, den **Energiebedarf pro Token** zu reduzieren.

Reduzierung der Eingabe: Prompts, auch Mega-Prompts, sollten Sie zwar mit ausreichend Informa-tionen versehen, doch so kurz wie möglich formulieren. Entfernen Sie Füllwörter, Wiederholungen und Platzhalter. Obwohl Kontext sinnvoll ist, sollten Sie damit maßvoll umgehen: Kürzen oder entfernen Sie übermäßige Beispiele oder lange Anweisungen, wenn sie nicht erforderlich sind.

Beispiel für Tokenbegrenzung beim Prompten

Das hier genannte Beispiel können Sie in den Einstellungen einiger Modelle fest-legen, oder Sie schreiben es zu dem Prompt direkt ins Chatfenster. Das Ausgabe-ergebnis wird je nach Varianz der „max_tokens“ länger oder kürzer ausfallen. So können Sie die Gen AI zwingen, weniger Energie für die Beantwortung auszugeben. Ebenfalls können Sie hier, z.B. bei Chat GPT, unter „model“ einstellen, welches GPT-Modell zur Beantwortung herangezogen werden soll. Durch die Wahl kleinerer Modelle kann man ebenfalls Energie sparen (vgl. Tipp 7).

```
{  
„model“: „gpt-4o-mini“,  
„prompt“: „Erkläre Green Prompting für KMU.“,  
„max_tokens“: 200  
}
```

Mega Prompt

Durch einen gut durchdachten Prompt können Sie sich viel Arbeitszeit, Rechenzeit und Energie sparen. Mehr Vorarbeit in einem langen Prompt mit viel Kontext zahlt sich aus: Die Antwortqualität ist sehr wahrscheinlich höher, es braucht weniger „Nachprompten“ um zu einem zufriedenstellenden Ergebnis zu kommen.

Tipp 6: Mögliche Struktur eines Megaprompts

Format: In welcher Form soll das Ergebnis ausgegeben werden?

Persona: Welche Rolle soll die Gen AI einnehmen? Hier kann man bspw. Berufsbezeichnungen mit anderen Merkmalen kombinieren (z.B. Geschlecht, Merkmal, Verortung/Hintergrund).

Aufgabe: Was genau soll getan werden?

Kontext: Welche Rahmenbedingungen und Einschränkungen müssen beachtet werden?

Schrittfolge: Wie soll die KI vorgehen? Hier kann man z.B. wählen ob „Reasoning“, d.h. energieintensives, tiefes Nachdenken verwendet werden soll oder ob andere Anweisungen zur Vorgehensweise gegeben werden (siehe Tipps zur Prompt Planung).

Ziel: Was soll erreicht werden? Wie soll das Ergebnis ausgegeben werden bzw. aufbereitet sein? (siehe Tipps zur Prompt Planung)

Solche gut vorbereiteten Mega Prompts sind umfangreiche und oft mehrstufige Eingaben, die darauf abzielen, besonders komplexe oder detaillierte Aufgaben zu lösen. Sie bestehen aus einer Kombination von Aufgabe(n), Anweisung(en) sowie Kontextinformationen, die der Gen AI helfen, umfassende, zusammenhängende Antworten zu generieren.

Beispiel für einen Mega Prompt oder ein grünes Praxisbeispiel?

```
:// definiere rolle: „Du bist ein Reiseführer.“
```

```
:/ setze tonalität: „Sei freundlich und enthusiastisch.“
```

```
:/ liefere kontext: „Eine Familie möchte einen Sommerurlaub in Europa machen und benötigt Empfehlungen.“
```

```
:/ beschreibe aufgabe: „Erstelle einen detaillierten 7-tägigen Reiseplan für einen Familienurlaub in Wien, Österreich. Gib Empfehlungen für familienfreundliche Aktivitäten, Speisemöglichkeiten und kulturelle Erlebnisse.“
```

```
:/ definiere ausgabeformat: „Erstelle eine Liste der möglichen Aktivitäten, sortiert nach Wochentag und Tageszeit.“
```

Passende Modellgröße wählen

Stellen Sie sich die Frage: In welchem Modell prompte ich überhaupt und warum? Hier gibt es großes Potenzial, Energie zu sparen

Nach aktuellem Forschungsstand ist der wichtigste Faktor in dem Energieaufwand, den eine Chat-Bot-Anfrage benötigt, die Wahl des KI-Modells. Vereinfacht gesagt: Kleinere Modelle brauchen weniger Rechenkapazität und daher ist der Ressourcenverbrauch geringer. Für bestimmte Anwendungen reicht dies vollkommen aus und sollte vor bzw. beim Prompten unbedingt mitgedacht werden!

Kleinere Sprachmodelle werden aufgrund der Verwendung von weniger Parametern als Large Language Models (LLM) auch **Small Language Models (SLM)** genannt.

Parameter sind interne Variablen, wie z. B. Gewichtungen und Verzerrungen, die ein Modell während des Trainings lernt und die Verhalten bzw. Leistung des Modells beeinflussen. **SLM** haben **Millionen bis Milliarden**, **LLM** dagegen **Hunderte von Milliarden bis Billionen** Parameter. Die Effizienz resultiert aus verschiedenen Techniken, z.B. aus der Wissensdestillation, bei der Wissen und Denkweise eines großen „Lehrermodells“ an das „Schülermodell“ / SLM weitergegeben werden.

Beispiele von SLM: DistilBERT, Gemma, GPT-4o mini, Granite, Llama, Ministral, Phi

Mehr Infos zu den SLM und ihrer Eignung:

<https://www.ibm.com/de-de/think/topics/small-language-models>

Mittlerweile gibt es im Unternehmenskontext den Trend, so genannte kleine Sprachmodelle (SLM) gezielt einzusetzen. Dahinter stecken nicht veraltete Mini-Modelle, sondern der Fokus auf effiziente, schnelle, spezialisierte und datenschutzkonforme Lösungen (mehr dazu im Kasten sowie im Tipp: Nachhaltige Infrastruktur nutzen).

Tipp 7: Small Language Models nutzen:

Wenn möglich, nutzen Sie kleinere oder spezialisierte Modelle (z. B. GPT-4 Turbo, Mini-Modelle oder lokale Lösungen).

Hier gilt nicht: „Viel hilft viel“, sondern: Wählen Sie das richtige Modells je nach Anwendungsfall.

Testen Sie das für sich oder Ihr Team aus und nehmen Sie sich auch zukünftig immer wieder einen Raum, um verschiedene Modelle wie z.B. neuere Mini-Modelle auch bei komplexeren Anforderungen auszutesten.

Entscheiden Sie sich für ein KI-Modell, dass für den jeweiligen Anwendungsfall passt, denn::

- ✓ Spezialisierte Modelle passen nicht für jede Anwendung (z.B. wählen Sie zur Bilderstellung ein multimodales Modell),
- ✓ Biases oder Halluzinationen können, je nach Modell und Fall auftreten (d.h. prüfen Sie die Ergebnisse auf ihre Güte),
- ✓ Sehr komplexe Prompts mit viel Kontext und hoher Spezifität könnten ein Modell überfordern.

In der Praxis ist es z.B. denkbar, dass Sie ein zwei oder drei Modelle zur Auswahl haben und diese je nach anstehender Aufgabe wählen. In diesem Fall könnten Sie für einfache Aufgaben ein SLM lokal betreiben (z.B. auf dem eigenen Rechner/Intranet ohne Anbindung ans Internet).

Welches Modell für welche Aufgabe?

Diese Entscheidungsmatrix gibt Orientierung für die Wahl der passenden Modellgröße, als „Daumen-Regel“.

| Kriterium | SLM (kleines Model) | LLM (großes Model) |
|--|--|---|
| Aufgabenkomplexität | Ist die Aufgabe eng und repetitiv (z.B. Datenextraktion)? | Erfordert sie breites Allgemeinwissen und Kreativität ? |
| Latenz & Echtzeitbedarf | Muss die Antwort sofort verfügbar sein, auch offline ? | Ist eine geringe Latenz tolerierbar ? |
| Datenschutz | Müssen sensible , proprietäre Daten lokal verbleiben? | Sind die Daten öffentlich oder kann eine externe Verarbeitung toleriert werden? |
| Kosten & Skalierung | Erwarten Sie viele Nutzer und hohe Skalierung ? | Sind die laufenden Kosten weniger kritisch? |
| Entwicklungshorizont & Teamskalierung | Hat Ihr Team das Know-how , ein Modell zu trainieren und zu warten ? | Bevorzugen Sie eine schnelle “Plug-and-Play” -Lösung? |

Für größere Fragen, z.B. bei der der Auswahl eines Modells in Skalierung (z.B. für eine ganze Abteilung, einen häufigen Anwendungsfall etc.), ist eine spezifische Auseinandersetzung mit den jeweiligen Modellen und ihren Stärken in Bezug auf den Anwendungsfall empfohlen.

Wip
gra
aies

Upgrades: Generative KI noch „grüner“ einsetzen!

Neben diesen Basics des Green Prompting klangen bereits in den Tipps verschiedene Möglichkeiten an, wie man Gen AI noch „grüner“ gestalten bzw. einsetzen kann. Da diese jedoch über das reine „Prompten“ hinausgeht, haben wir sie hier in diesem Kapitel gesammelt.

Tipp 8: Nutzen Sie die Guidelines for Green AI

Der Green AI Hub hat Leitlinien herausgegeben bzw. aktualisiert diese zu den wichtigsten Feldern rund um den Aufbau und Betrieb von nachhaltigerer KI:

vgl. GreenAI Hub; 2022, [LINK](#)

Nachhaltige Infrastruktur

Mit Bezug auf das Betreibermodell, das Energiemanagement vor Ort sowie bei genutzter Hardware und Software gibt es Potenziale, Gen AI nachhaltiger zu betreiben.

Lokales Betreiben einer sparsamen Infrastruktur

Gen AI kann man grundsätzlich immer vor Ort betreiben. Große Sprachmodelle bedürfen jedoch auch großer Rechenkapazitäten. Um eine angemessene Qualität der Ergebnisse zu erreichen, gibt es die Möglichkeit, einen kleinen, effizienten Chatbot (SLM / vgl. Tipp: Passende Modellgröße wählen) mit den eigenen Unternehmensinformationen anzureichern. Dies ist eine datenschutzkonforme, zuverlässige, energieeffiziente und in der Praxis gängig genutzte Möglichkeit:

Die Kombination einer Gen AI mit der Retrieval-Augmented Generation (RAG)-Technologie verzichtet auf aufwendiges Fine-Tuning des Chatbots, sondern zieht relevante Unternehmensdaten bspw. aus einer Vektordatenbank als Wissen

heran. Das Sprachmodell basiert die Antworten auf diesen Informationen, ohne sie selbst vorzuhalten. So können Sie mit eigenen Dokumenten „chatten“, die Rechenlast bleibt gering und die Energiebilanz positiv.

Mehr Informationen zur technischen Umsetzung finden Sie z.B. hier: <https://www.digitalzentrum-fokus-mensch.de/kos/WNetz?art=News.show&id=2850>

Größenvergleich von Sprachmodellen

Eine Studie der International Energy Agency von 2025 beschreibt Ergebnisse aus Experimenten, bei denen folgende Vergleichswerte gemessen wurden:

Derselbe Prompt resultierte z.B. in einem sehr kleinen SLM (SmolLM2-1.7B-Instruct) für 0,1 Wh direkten Stromverbrauch des Grafikprozessors am PC (GPU). Ein mittleres Sprachmodell verbrauchte 4 Wh, d.h. 40 mal so viel Strom (Llama 3.3). Das große LLM benötigte sogar 9 Wh (DeepSeek-R1).

vgl: International Energy Agency, (2025): Energy and AI, S.45ff.

KI Anbieter und ihr „Energy Scoring“

Zum Zeitpunkt der Erstellung dieses Guides existierten keine Reporting- und Dokumentationsstandards bezüglich des Ressourcenverbrauchs von KI-Betreibern, obgleich die Forderungen in der Gesellschaft lauter und auch verstärkt auf nationaler Ebene wie z.B. durch das Umweltbundesamt bewertet werden. Bis mehr Transparenz herrscht, kann man als Nutzer*in lediglich auf Informationen Dritter zugreifen.

Bei **der direkten Beschaffung von Rechenleistung** können Sie auf zertifizierte Rechenzentren zurückgreifen oder sogar eigene Rechenkapazitäten zertifizieren lassen. Der TÜV bietet

Zertifikate mit Fokus auf physische Sicherheit, mit der Möglichkeit ein Zusatzzertifikat für Energieeffizienz zu erlangen („PUE“) . Für Rechenzentren gibt es verschiedene Zertifizierungen und Umweltzeichen, wie z.B. Blauer Engel , ISO 50001, DIN EN 50600-4-2 oder das Siegel „Sustainable Data Center“ vom TÜV Rheinland. Sie können bei der Beschaffung außerdem beachten, wie z.B. sich das Energiemanagement bzw. die Kühlung oder Wasserbeschaffung eines Rechenzentrums auswirken oder über den Zeitpunkt, wann Rechenleistung bezogen wird, die eigenen Prozesse bezüglich des CO₂Fußabdrucks optimieren.

Tipp 9: Infos zum Energieverbrauch von Gen AI

Aktuell bieten Projekte wie z.B. die Kooperation von Salesforce & Hugging Face, das AI Energy Scoring, eine grobe Orientierung. Hier werden die Sprachmodelle in ihren Einsatzbereichen z.B. Textgenerierung oder Videogenerierung nach Energieverbrauch gerankt.

siehe: Hugging Face (2026): AI Energy Score, [LINK](#)

Tipp 10: Technik green & easy nutzen

Nutzen Sie vorhandene Hardware lange, ggf. reparieren bzw. wiederaufbereiten Sie diese (Upgrades auf aktuelle Bedarfe, möglich bei modularen Systemen).

Bei Neukauf, wählen Sie modulare Hardware (Reparierbarkeit, Skalierbarkeit, Anpassungsfähigkeit), achten Sie auf Energieeffizienz (z.B. wiederaufbereitete Produkte, hohe Energieeffizienzklasse) und bevorzugen Sie umweltzertifizierte Produkten (z.B. Blauer Engel).

Green IT / Hardware

Die Auswahl ressourcenschonender Hardware ist ein zentraler Hebel zur Senkung des Energieverbrauchs bei der KI-Nutzung. Mit Bezug auf die Hardware bzw. Strategien rund um Green IT können Sie mit Blick auf Serverinfrastrukturen, Kühlsysteme etc. auch sehr viele Potentiale erschließen.

Grüne Software (mit KI Anbindung)

Neben der Hardware ist auch die Software bedeutsam für einen nachhaltigeren Betrieb von Gen AI. Wenn Software auf Basis eines SLM läuft (vgl. Tipp 7) und auf seine Anwendungsfälle optimiert ist, ist dies energieeffizienter, als Software auf Basis eines nicht fine-getunten, generischen LLMs zu nutzen. Prüfen Sie, ob eine Umsetzung der Aufgaben mit Edge Computing mit Hilfe einer lokal laufenden Software möglich ist.

Energiemanagement der eigenen Strukturen (PV, Wasser, Ökostrom etc.)

Egal, ob man eine KI lokal betreibt oder lediglich Anbieter nutzt, man verbraucht natürlich auch an den eigenen Standorten Ressourcen. Denken Sie daran, den Energieverbrauchs ihrer bestehenden IT-Infrastruktur zu analysieren, um weitere Einsparpotenziale zu erkennen:



Server-Infrastrukturen



Speicherplattformen



Softwareanwendungen



Netzwerke



Hardware



Energiebereitstellung
(Stromanbieter, Energiequellen)

Energieberater*innen, IT-Verantwortliche und Entscheider*innen können hier gemeinsam die Energieeffizienz erhöhen und somit auch eine umweltschonendere Grundlage für die Gen AI Nutzung schaffen.

Organisation und Zusammenarbeit

Eine nachhaltige Infrastruktur bringt nur so viel, wie sie genutzt wird. Und auch in der Art der Nutzung steckt sehr viel Potenzial, Ressourcen zu sparen und die Umwelt zu schonen. Eine Unternehmenskultur, in der Nachhaltigkeit hoch priorisiert und gelebt wird, ist – aus Sicht umweltfreundlicher KI-Nutzung – die Zielmarke. Doch auch in anderen Fällen können kleine Weichenstellungen viel bringen.



Den Beitrag zu Meeting Games finden Sie unter folgendem Link oder Sie folgen dem QR-Code.

<https://www.digitalzentrum-zukunftskultur.de/angebot/meeting-games-14861/>

Sensibilisieren

Vielen Menschen ist nicht klar, welche Umweltauswirkungen und Ressourcenkosten der Einsatz von Gen AI hat. Doch wenn sich Menschen hilflos fühlen, ändern sie ihr Verhalten wahrscheinlich nicht. Das Verknüpfen mit konkreten Handlungsmöglichkeiten wie z.B. Green Prompting ist daher empfehlenswert. Dies kann über verschiedene Wege erfolgen, z.B. Infoschreiben, Team Meetings, Schulungen, KI-Arbeitskreise oder in gezielten Veranstaltungen.

Gemeinsam besser werden

Der Austausch im Team bzw. mit Personen aus ähnlichen Tätigkeitsfeldern ist wertvoll. Welche Prompts waren zielführend, wo sind wir gescheitert, welche Erfahrungen wurden mit verschiedenen Modellen oder Anwendungen gemacht? Speichern Sie Portfolios „guter Beispiele“ so, dass sie schnell und einfach nutzbar sind, idealerweise von mehreren Personen oder gar Abteilungen. Green Prompting ist ein Lernprozess, kein fester Standard.

Leitplanken und Regeln

Gemeinsam festgelegte Regeln und Leit- oder Richtlinien helfen allen Menschen in Unternehmen bei der Orientierung, wie sie Gen AI nutzen können und dürfen. Die KI-Verordnung bzw. Gesetzgeber wird hierzu verbindliche Regeln für KI-Betreiber bzw. Anwender festlegen. Diese dienen nicht vorrangig dem Umweltschutz, sondern sind z.B. für eine Risikominimierung, Datenschutz oder IT-Sicherheit wichtig.



Unsere Publikation zur KI Verordnung
finden Sie unter folgendem Link oder
Sie folgen dem QR-Code.

[https://www.digitalzentrum-zukunftskultur.de/
wp-content/uploads/2026/01/KI-VO-
Leitfaden.pdf](https://www.digitalzentrum-zukunftskultur.de/wp-content/uploads/2026/01/KI-VO-Leitfaden.pdf)

Die Übersetzung dieser Regeln in die Unternehmensprozesse z.B. durch KI-Nutzungsrichtlinien oder Betriebsvereinbarungen können Sie gut nutzen, um Green Prompting im eigenen Unternehmen zu etablieren.

Auf einen Blick



Auf einen Blick zusammengefasst finden Sie unsere Green Prompting Guidelines auf einem Poster, das nur einen Klick entfernt ist.

<https://www.digitalzentrum-zukunftskultur.de/material/green-prompting-guidelines-15255/>

Mittelstand-Digital Zentrum Zukunftskultur

In Kooperation mit
Zukunftszentrum Berlin
Digitale Arbeitswelten gestalten

Green Prompting Guidelines - Auf einen Blick

Basics: Diese Tipps machen Dein Prompten "grüner"

- + Kein Prompt ist der grünste Prompt
- + Behalte Dein Ziel im Auge
- + Prompts präzise und konkret
- + Gib Grenzen und Formate vor
- + Formuliere Ziele und Rolle
- + Mega-Prompts bringen Dich schneller zum Ziel:
Format, Rolle, Aufgabe, Kontext, Schrittfolge und Ziel
- + Stelle eine Tokenbegrenzung ein
"model": "model-XYZ-mini",
"prompt": "Erkläre Green Prompting für KMU.",
"max_tokens": 200
- + Prüfe bzw. nutze ein Small Language Model
z.B. Mini-Modelle oder eine lokale Lösung

Upgrade: So kannst Du KI noch "grüner" einsetzen

- + **Nutze nachhaltige KI-Infrastruktur**
Lokale KI, sparsamer Anbieter, Green IT (Hard- & Software), Energiemanagement (PV, Wasser, Okeostrom etc.)
- + **Organisation und Zusammenarbeit**
Sensibilisiere Dein Team, werdet gemeinsam besser, nutzt Leitplanken und Regeln

© Clutch Illustrations in Canva Pro

Ausblick

Wir hoffen mit diesem Guide eine Perspektive zu eröffnen, um eine immer noch verhältnismäßig neue Technologie effektiv und effizient zu nutzen und Nachhaltigkeit mit positiver Unternehmensentwicklung zu kombinieren. Die Diskussionen um die ökologischen Folgeerscheinungen Generativer KI insbesondere hinsichtlich ihres Ressourcen- und Energiebedarfs und der verbundenen Effekte auf Umwelt und Mensch steckt noch in den Kinderschuhen. Weder gibt es aktuell ausreichend belastbare Zahlen noch messbare Indikatoren und entsprechend benötigte Bereitschaft der Anbieter, die

Umweltbelastungen durch die Nutzung ihrer Anwendungen transparent zu machen. Wenn wir mit dem Green Prompting Guide etwas Starthilfe für bewussteres und effektiveres Prompten geben konnten, ist unser Ziel schon erreicht.

Zukunft

Ausblick

Die Zukunft des Themas, der Technologieentwicklung und -nutzung ist schwer zu prognostizieren. Aus Fachkreisen hören wir, dass in den kommenden Jahren zwei Trends zu erwarten sind:

Die Nutzung kleinerer, spezialisierter Anwendungen (auch lokal) wird zunehmen, da dies für viele Anwendungsfälle die beste Kosten-Nutzen-Bilanz aufzeigt. Andererseits wird vermutlich auch die Größe der marktführenden großen Sprachmodelle wachsen, um ihr „Weltwissen“, die Qualität ihrer Ergebnisse bei einer großen Breite von Aufgaben zu erhöhen.

Ebenfalls rechnen wir damit, dass es mehr Transparenz für KI-Infrastruktur, -Training und Nutzung geben wird. Dies wird dazu führen, dass auch der Bereich Green Prompting besser erforscht und erschlossen werden kann.

Wir hoffen, dass dies genutzt werden wird, um eine unnötige Übernutzung zu verhindern und Reboundeffekte nicht zu Lasten der Umwelt fallen zu lassen.

Zarte Pflänzchen: Transparenz am Beispiel MLFlow

Mit Blick auf die Sichtbarkeit von Emissionen auf Prompt- und Modellebene entstehen aktuell Tools, mit denen künftig mehr Transparenz erreicht werden kann. Z.B. ermöglicht das Open-Source-Dashboard MLflow GenAI KI-Applikationen zu entwickeln und, z.B. in Kombination mit der Anwendung CodeCarbon Metriken wie CPU, GPU, Arbeitsspeicher, Festplatten- und Netzwerkauslastung zu evaluieren. Expert*innen können so näherungsweise Emissionen und Energieverbräuche messen. Derzeit gibt es jedoch noch keine „Plug-and-Play-Lösung“ für Nutzende. Wir hoffen jedoch, dass sich dies bald ändern wird und Sie im Rahmen von KI-Dashboards die Performance ihrer Prompts oder KI-Agenten inkl. der Energieverbräuche live sehen können.

vgl: Vanderschueren, Martijn (2026) Tracking Carbon Emission During MLOps [LINK](#)



WWeit

er

ver

tie

fen

Literatur

- ABV Creative, Medium Magazin (2024): Doomprompting Is the New Doomscrolling — And It's Wrecking Your Work, <https://abvcreative.medium.com/doomprompting-is-the-new-doomscrolling-and-its-wrecking-your-work-cf69e514772e>
- Adamska, M. et al. (2025): Green Prompting, Lancaster, arXiv preprint, <https://arxiv.org/abs/2503.10666>
- Bergman, D. (2025): Was ist Wissensdestillation? (IBM Think Artikel). <https://www.ibm.com/de-de/think/topics/knowledge-distillation>
- Caballar, R.M. (2025): Was sind kleine Sprachmodelle (IBM Think Artikel)? <https://www.ibm.com/de-de/think/topics/small-language-models>
- Cross, G. (2025): Doomprompting: Endless tinkering with AI outputs can cripple IT results (Artikel). <https://www.cio.com/article/4056586/doomprompting-endless-tinkering-with-ai-outputs-can-cripple-it-results.html>
- Dauner, M., Socher, G. (2025): Energy costs of communicating with AI., in: Front. Commun., Sec. Science and Environmental Communication Vol. 10: 2025, München. <https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1572947/full>
- DeMartino, V. et al. (2025): Green Prompt Engineering: Investigating the Energy Impact of Prompt Design in Software Engineering, Padova, arXiv preprint. <https://arxiv.org/pdf/2509.22320>
- Gailhofer, P. (2025): Bessere Spielregeln für digitale Berater. Chatbots als universelle Experten oder als „epistemisches Risiko“ für die Nachhaltigkeitstransformation?, Policy Brief, Oeko-Institut, Berlin. <https://www.oeko.de/publikation/bessere-spielregeln-fuer-digitale-berater/>
- Gailhofer, P. et al. (2025): Umweltrechtliches Regulierungskonzept für algorithmenbasierte Entscheidungssysteme. Umweltbundesamt, Dessau-Roßlau. <https://doi.org/10.60810/openumwelt-8126>
- Gröger, J. et al. (2025): Umweltauswirkungen Künstlicher Intelligenz, Greenpeace e.V. & Öko-Institut GmbH, Hamburg, <https://www.greenpeace.de/publikationen/20250514-greenpeace-studie-umweltauswirkungen-ki.pdf>
- Guidelines for Green AI, Green AI Hub, Berlin, <https://www.green-ai-hub.de/ki-fuer-ressourceneffizienz/guidelines-for-green-ai>
- Hochschule München (2025): Neue Studie zeigt: Je besser KI antwortet, desto größer ihr ökologischer Fußabdruck (Pressemitteilung). https://hm.edu/aktuelles/news/news_detailseite_458176.de.html

- Honroth, T. et al. (2024): Retrieval Augmented Generation (RAG): Chatten mit den eigenen Daten (Artikel). <https://www.iese.fraunhofer.de/blog/retrieval-augmented-generation-rag/>
- Hugging Face (2025): AI Leaderboard (Website). <https://huggingface.co/spaces/AIEnergyScore/Leaderboard>
- International Energy Agency (2024): Energy and AI, Paris. <https://iea.blob.core.windows.net/assets/dd7c2387-2f60-4b60-8c5f-6563b6aa1e4c/EnergyandAI.pdf>
- JazJon (2022): The AI prompt was “Salmon in the river”, Reddit. https://www.reddit.com/r/aiArt/comments/ybnj7n/the_ai_prompt_was_salmon_in_the_river/
- Johnson, J. (2025): Small Language Models (SLM): A Comprehensive Overview (Hugging Face Artikel). <https://huggingface.co/blog/jjokah/small-language-model>
- Ligozat, A.-L. et al. (2022): Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. Sustainability 2022, 14, 5172. <https://doi.org/10.3390/su14095172>
- Stobbe, L. et al. (2025): Strombedarf und Carbon Footprint der IKT in Deutschland 2010 – 2036, Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration & Kompetenzzentrum »Green ICT @ FMD« der Forschungsfabrik Mikroelektronik Deutschland, Berlin. <https://elektronikforschung.de/service/publikationen/strombedarf-und-carbon-footprint-der-ikt-in-deutschland-2010-2036>
- Stojkovic, J. et al. (2024): Towards Greener LLMs: Bringing Energy-Efficiency to the Forefront of LLM Inference, Urbana /Champaign, arXiv preprint, <https://arxiv.org/abs/2403.20306>
- TÜV Nord (2025): Zertifizierung von Rechenzentren nach TSI, EN 50600 oder ISO/IEC 22237 (Website). <https://www.tuev-nord.de/de/dienstleistungen/auditierung-und-zertifizierung/zertifizierung-von-rechenzentren/>
- Vanderschueren, Martijn (2026) Tracking Carbon Emission During MLOps, <https://dataroots.io/blog/tracking-carbon-emission-during-mlops>
- Wallner, M. (2025): Grüne IT durch energieeffizientes Computing (Fachartikel). <https://www.it-administrator.de/Green-IT-durch-energieeffizientes-Computing>
- Winter, L., & Züger, T. (2025): Blind spot sustainability: Making AI's environmental impact measurable. In Digital Society Blog. Zenodo. <https://doi.org/10.5281/zenodo.16608366>
- Zhenya, J., Ming, J. (2026): A systematic review of electricity demand for large language models: evaluations, challenges, and solutions, in: Renewable and Sustainable Energy Reviews 2026:225. <https://www.sciencedirect.com/science/article/pii/S1364032125008329?via%3Dihub>

Kooperation mit dem Zukunftszentrum Berlin

Der Green Prompting Guide ist in Zusammenarbeit mit Kolleginnen des Zukunftszentrum Berlin entstanden.

Die Zukunftszentren:

Mit dem durch den Europäischen Sozialfonds Plus (ESF Plus) geförderten Programm strebt das Bundesministerium für Arbeit und Soziales (BMAS) an, Unternehmen, Arbeitnehmer und (Solo-)Selbstständige zu befähigen, den Wandel der Arbeit aktiv mitzugestalten. Die 14 regionalen und konsortialpartnerschaftlich umgesetzten Zukunftszentren beraten und vernetzen KMU, sie entwickeln und prüfen innovative Weiterbildungskonzepte direkt im Betrieb. Ein zentrales Merkmal dabei ist die partizipative Ausgestaltung von Einführungs- und Anwendungsprozessen für digitale Technologien und KI-basierte Systeme in enger Zusammenarbeit mit den Beschäftigten.

Weitere Informationen finden Sie unter: <https://zukunftszentren.de/>

Das Zukunftszentrum Berlin

Das „Zukunftszentrum Berlin – digitale Arbeitswelten gestalten“ (Zukunftszentrum Berlin) unterstützt Betriebe als Informations-, Beratungs- und Vermittlungsstelle bei der Digitalisierung und der Einführung von KI-Anwendungen. Es möchte die Potenziale von digitalen Anwendungen hervorheben und Lösungswege zur Überwindung möglicher Einführungshemmnisse aufzeigen. Die entwickelten Angebote (Beratung, Vermittlung sowie betriebliche Weiterbildungsangebote) richten sich an drei betriebliche Hauptzielgruppen: Führungskräfte/Personalverantwortliche, Beschäftigte und Betriebsrät*innen/ Interessensvertretungen und sind branchenoffen.

Mehr dazu: <https://zukunftszentrum-berlin.de/>

Ihre Ansprechpartnerinnen



Wiebke Reyels & Francesca Hermani

Zukunftszentrum Berlin

E-Mail: info@zukunftszentrum-berlin.de

Telefon: +49 (0)30 / 417 49 86 - 31



Das Projekt „Zukunftszentrum Berlin“ wird im Rahmen des Programms „Zukunftszentren“ durch das Bundesministerium für Arbeit und Soziales gefördert und von der Europäischen Union über den Europäischen Sozialfonds Plus (ESF Plus) sowie von der Senatsverwaltung für Arbeit, Soziales, Gleichstellung, Integration, Vielfalt und Antidiskriminierung aus Mitteln des Landes Berlin kofinanziert.

Gefördert durch:



**Zukunftszentrum
Berlin**
Digitale Arbeitswelten gestalten



Mittelstand-Digital

Das Mittelstand-Digital Netzwerk bietet mit den Mittelstand-Digital Zentren und der Initiative IT-Sicherheit in der Wirtschaft umfassende Unterstützung bei der Digitalisierung mit dem Schwerpunkt Künstliche Intelligenz. Kleine und mittlere Unternehmen profitieren von konkreten Praxisbeispielen und passgenauen, anbieterneutralen Angeboten zur Qualifikation und IT-Sicherheit. Das Bundesministerium für Wirtschaft und Energie ermöglicht die kostenfreie Nutzung der Angebote von Mittelstand-Digital.

Weitere Informationen finden Sie unter: www.mittelstand-digital.de.

Das Mittelstand-Digital Zentrum Zukunftskultur

Unser Team unterstützt Ihr Unternehmen, eine zukunftsorientierte Unternehmenskultur aufzubauen. Unser Weg zum Ziel geht dabei über neue Potenziale der Künstlichen Intelligenz, KI Readiness, Möglichkeiten der Digitalisierung, die Nutzung neuer Plattformen und eine nachhaltige Entwicklung Ihrer Geschäftsprozesse. Ein systematisches Veränderungsmanagement, wertschätzende Kommunikation und die Fokussierung der eigenen digitalen Verantwortung sind dabei Schlüssel zu langfristigem Erfolg. Über unsere Veranstaltungen, Medienformate und Praxisprojekte vermitteln wir Ihnen dazu wertvolle Impulse und praxisnahes Know-how:

www.digitalzentrum-zukunftskultur.de

Ihre Ansprechpartner*innen



Nachhaltigkeit und Digitalisierung

Susann Feuerschütz

E-Mail: susann.feuerschuetz@hnee.de

Telefon: +49 (0)3334 / 657 - 341

Mittelstand-Digital Zentrum Zukunftskultur



Nachhaltigkeit und Digitalisierung

Gerrit Neuhaus

E-Mail: gerrit.neuhaus@hnee.de

Telefon: 03334 / 657 -436

Mittelstand-Digital Zentrum Zukunftskultur

Weitere kostenlose Angebote finden Sie auf:
www.digitalzentrum-zukunftskultur.de