



Wie sag ich's der KI?

Praxisnahe Techniken für erfolgreiches Prompting von KI-Chatbots

René Perfözl

Freie Universität Berlin, FUB-IT, Abt. Lehre, Studium und Forschung
Arbeitsbereich E-Learning und E-Examinations (EEE)

VERANSTALTET VON:



IM RAHMEN EINES PROJEKTES VON:



GEFÖRDERT VON:



Inhalt

1. Einführung

2. Modell und Datenbasis

3. Prompt Design

4. Praxisphase

5. Q&A

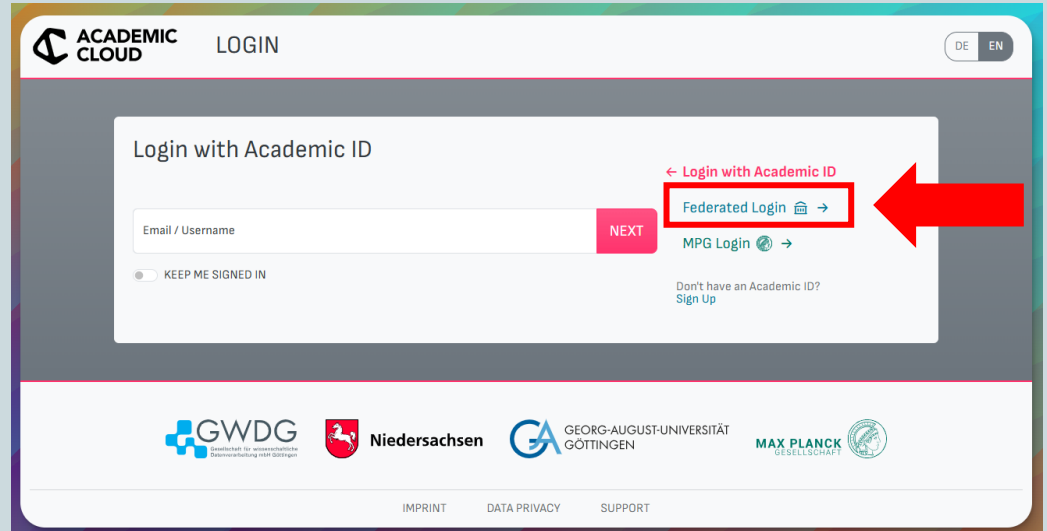
„ChatAI“ – KI-Chatbot

- **ChatAI:** Angebot von GWDG/KISSKI (KI-Servicezentrum gefördert vom BMBF)
- **hoher Datenschutzstandard:** Eingaben werden bei den selbstgehosteten Modellen nicht auf den Servern gespeichert
- Hinweis: Bitte **keine personenbezogenen** oder **urheberrechtlich geschützten** Daten eingeben!
- Bei anderen Anbietern: **Training ausstellen**

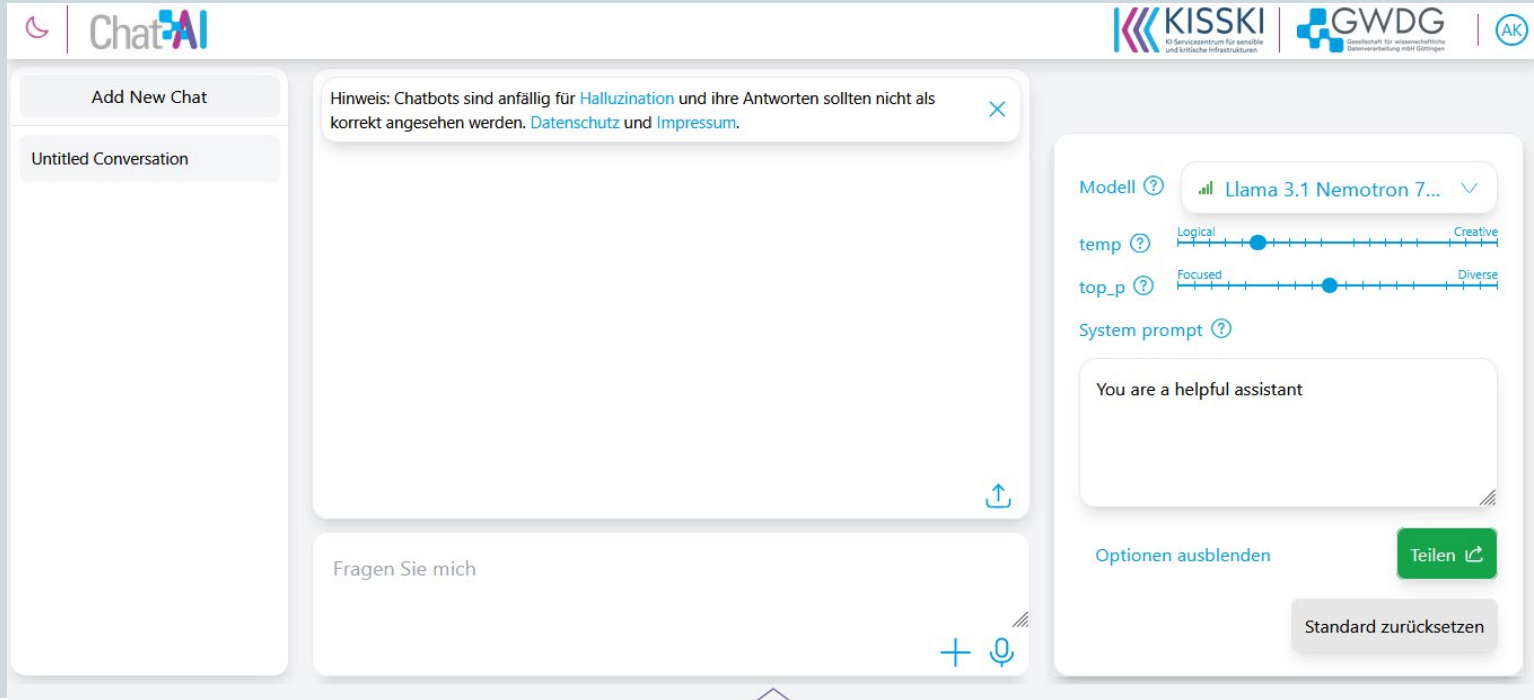
„ChatAI“ – Zugriff mit Uni-Account

Föderierter Zugang mit **eigenem Universitätsaccount** (via Academic Cloud):

<https://chat-ai.academiccloud.de/>



„ChatAI“ - Benutzeroberfläche



„ChatAI“ - Modell

The screenshot displays the ChatAI web interface. On the left, there's a sidebar with 'Add New Chat' and 'Untitled Conversation'. The main area features a warning banner about hallucinations, followed by 'Aktuelle Modell-Empfehlungen:' (Current Model Recommendations:). A red arrow points from this text to the 'Modell' dropdown menu on the right, which is currently set to 'Llama 3.1 Nemotron 7...'. Below the model selection, there are sliders for 'temp' (ranging from Logical to Creative) and 'top_p' (ranging from Focused to Diverse). A 'System prompt' field contains the text 'You are a helpful assistant'. At the bottom right, there are buttons for 'Optionen ausblenden' (Hide options), 'Teilen' (Share), and 'Standard zurücksetzen' (Reset to standard).

ChatAI

Add New Chat

Untitled Conversation

Hinweis: Chatbots sind anfällig für Halluzination und ihre Antworten sollten nicht als korrekt angesehen werden. [Datenschutz](#) und [Impressum](#).

Aktuelle Modell-Empfehlungen:

- **Mistral Large Instruct**
- **Llama 3.1 Nemotron 70B Instruct**

Fragen Sie mich

Modell ? Llama 3.1 Nemotron 7... ▾

temp ? Logical Creative

top_p ? Focused Diverse

System prompt ?

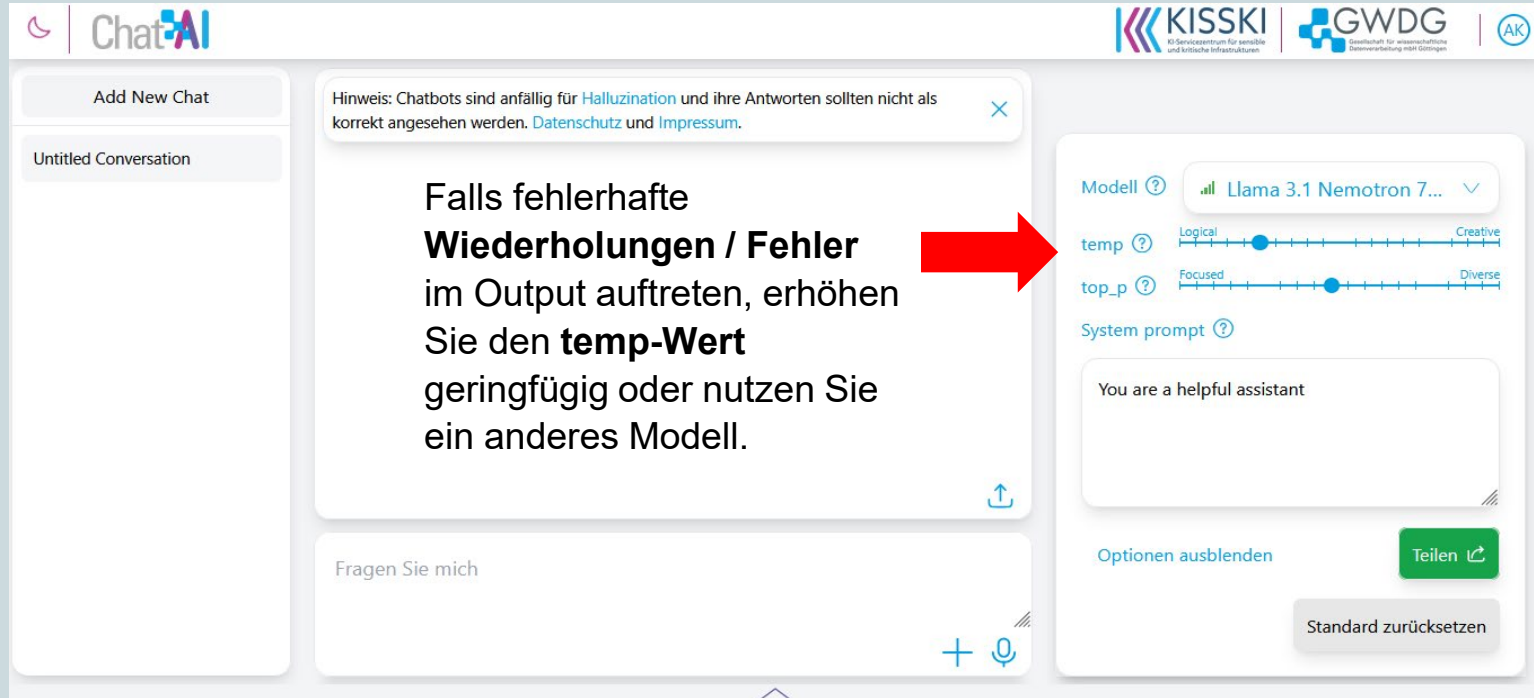
You are a helpful assistant

Optionen ausblenden

Teilen ↗

Standard zurücksetzen

„ChatAI“ - Temperatur



The screenshot shows the ChatAI interface. On the left is a sidebar with 'Add New Chat' and 'Untitled Conversation'. The main area contains a warning box: 'Hinweis: Chatbots sind anfällig für Halluzination und ihre Antworten sollten nicht als korrekt angesehen werden. Datenschutz und Impressum.' Below this is a text box with the instruction: 'Falls fehlerhafte **Wiederholungen / Fehler** im Output auftreten, erhöhen Sie den **temp-Wert** geringfügig oder nutzen Sie ein anderes Modell.' A red arrow points from this text to the 'temp' slider in the settings panel on the right. The settings panel shows 'Modell' set to 'Llama 3.1 Nemotron 7...', 'temp' on a 'Logical' to 'Creative' scale, and 'top_p' on a 'Focused' to 'Diverse' scale. Other options include 'System prompt' (set to 'You are a helpful assistant'), 'Optionen ausblenden', 'Teilen', and 'Standard zurücksetzen'.

ChatAI

Add New Chat

Untitled Conversation

Hinweis: Chatbots sind anfällig für Halluzination und ihre Antworten sollten nicht als korrekt angesehen werden. [Datenschutz](#) und [Impressum](#).

Falls fehlerhafte **Wiederholungen / Fehler** im Output auftreten, erhöhen Sie den **temp-Wert** geringfügig oder nutzen Sie ein anderes Modell.

Modell ? Llama 3.1 Nemotron 7... ▾

temp ? Logical Creative

top_p ? Focused Diverse

System prompt ?

You are a helpful assistant

Optionen ausblenden

Teilen ↗

Standard zurücksetzen

Aufgabe

Schreiben Sie einen Prompt für eine typische Aufgabe in Ihrem Arbeitsbereich und lassen Sie sich Ergebnisse erstellen, z.B.:

- Lehrvorbereitung (Lernziele, Seminarplan, Lehrmaterialien)
- Quizfragen (unterschiedliche Kompetenzbereiche und Anforderungsniveaus)
- Projektplanung (Ziele, Projektstrukturplan, Kosten)
- Abstract zu CfP erstellen
- ...

Sie haben dafür **5 Minuten** Zeit.



Quelle: DALL·E 2025-03-02
09.49.24 - A simple and clean
illustration of a teacher giving a
task to a chatbot.

Einführung

Funktionsweise LLM und Ergebnisqualität

Funktionsweise

Sprachmodell = künstliches neuronales Netz,

- das basierend auf „erlernten“ statistischen Mustern in umfangreichen **Trainingsdaten**
- zu vorgegebenen **Anweisungen/Fragen** (= Prompts)

Antworten (Wort-/Satzabfolgen) vorhersagt.

„Der Stochastische Papagei“



Quelle: Bing Image Creator (2023):
futuristic machine parrot, digital art,
isometric, 4k

Ergebnisqualität

- Wichtigster Faktor: **Modell** und ggf. **Modell-Größe**
- Zugriff auf **hochwertige Daten** (im Eingabefenster, als Dokument oder per RAG-Ansatz)
- Gestaltung der Anweisung/Frage = **Prompt Design**

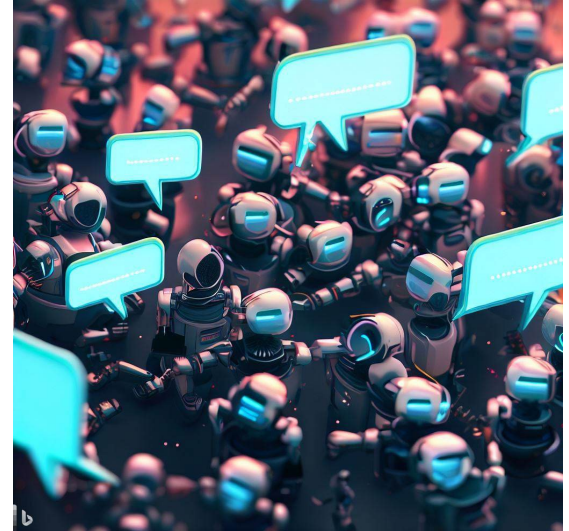


Quelle: DALL·E 2024-11-29 00.01.35
- Scientists and researchers making
adjustments to a chatbot

Modelle und Datenbasis

Aktuelle Modelltypen (GenKI)

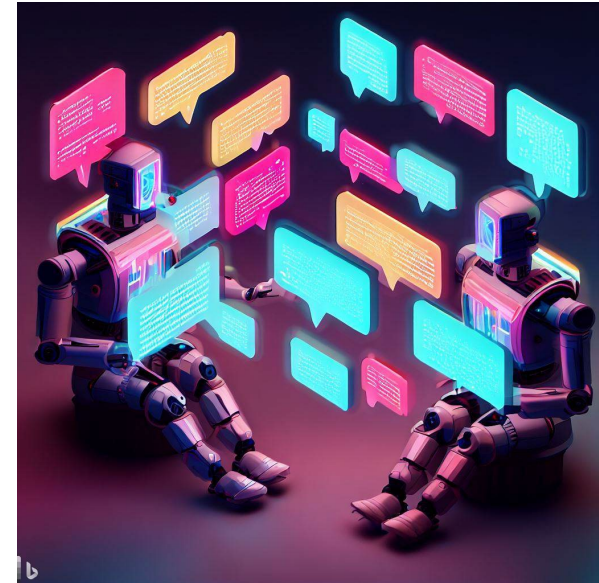
- **Multimodale Modelle** (z. B. *GPT-4o*, *Anthropic Claude Sonnet*): besonders für allgemeine Anfragen
- **Reasoning-Modelle** (z. B. *OpenAI o1*, *DeepSeek R1*): logische, arithmetische Fragestellungen (STEM-Bereich)
- **Spezialisierte Modelle:** Code-Modelle wie *OpenAI Codex*, Bildmodelle wie *DALL·E* oder *Midjourney*



Quelle: Bing Image Creator (2023): AI chatbots going crazy, digital art, very high quality, 8K, isometric

Modellgröße

- **je größer, desto leistungsfähiger (?)**
 - Llama 3.3 **70B** Instruct > Llama 3.1 **8B** Instruct
- kleinere Modelle können ebenso leistungsfähige Ergebnisse erzielen
 - mit passenden Datenquellen
 - gezieltem Finetuning



Quelle: Bing Image Creator (2023): Two AI chatbots answering different prompt, digital art, very high quality, 8K, isometric

Modellauswahl

- **allgemeine Benchmarks:** HELM, MMLU, BIG-Bench, SimpleQA o.ä.
- **eigene Tests** unerlässlich, z.B. mit KI-Kompass oder Chatbot-Arena (LMSYS-Corp.)

⇒ **Aufgabe:** Vergleichen Sie die Ergebnisse zweier unterschiedlicher Modelle beim KI-Kompass für denselben Prompt: z.B. DeepSeek R1 und GPT4o.

Sie haben dafür **5 Minuten** Zeit.



Quelle: Photo by Jakub Žerdzicki on Unsplash

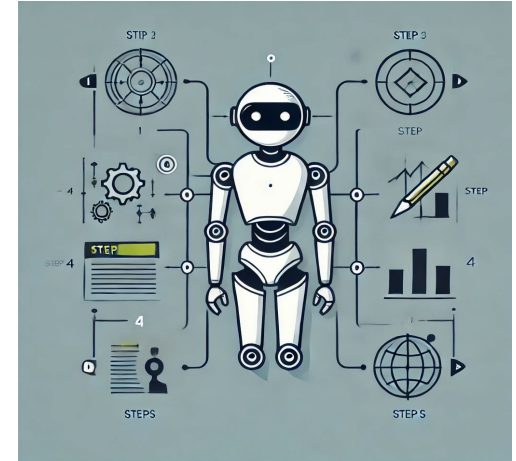
Datenbasis

Datenbasis	Vorteile	Nachteile
Wissen des Modells	• breites Allgemeinwissen	• fehlende/falsche Quellennachweise • fehlendes Fachwissen • keine aktuellen Informationen
Internet	• aktuelle Informationen • Quellennachweise • Weniger Halluzinationen	• begrenzte Auswahl an Quellen • Qualität und Glaubwürdigkeit
Eigene Daten	• präzisere Antworten • spezifisches Fachwissen • Quellennachweise (je nach Tool) • Weniger Halluzinationen	• begrenzte Auswahl an Quellen

Prompt Design

Allgemeine Hinweise

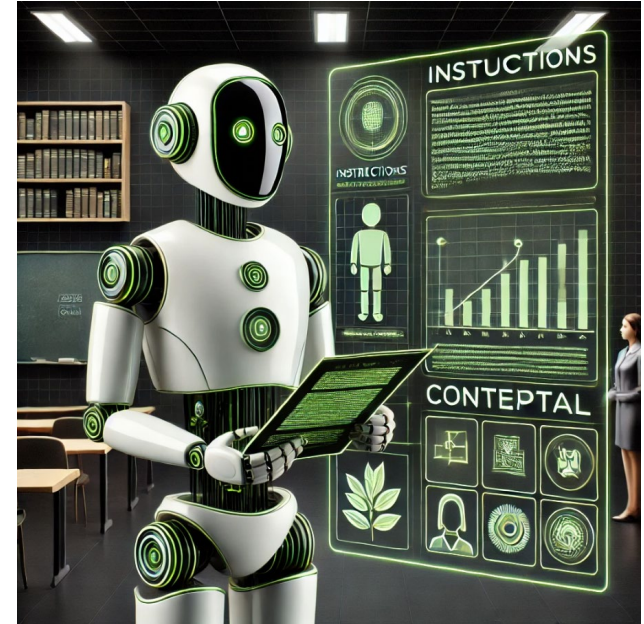
- **Iteratives** Vorgehen: testen, analysieren, anpassen
- **Chatverlauf beeinflusst** Folge-Antworten
 - „Vergessen“ → Vorgaben erneut eingeben
 - Themenwechsel → neuer Chat
 - Fehlende oder falsche Antworten → Prompt anpassen
- **Komponenten:** Aufgabe, Kontext, formale Vorgaben, Beispiele



Quelle: DALL-E 2024-11-28 14.05.28 - A simple and graphic representation of a chatbot figure structuredly solving a task in four steps

1. Aufgabe / Frage

- **Präzise** und **eindeutig** formulieren
- komplexe Aufgaben in **Unteraufgaben** aufteilen und einzeln prompten (*Chained Prompting*)
 - Intervention, falls unerwünschte Ausgaben
 - Roter Faden bei längeren Outputs



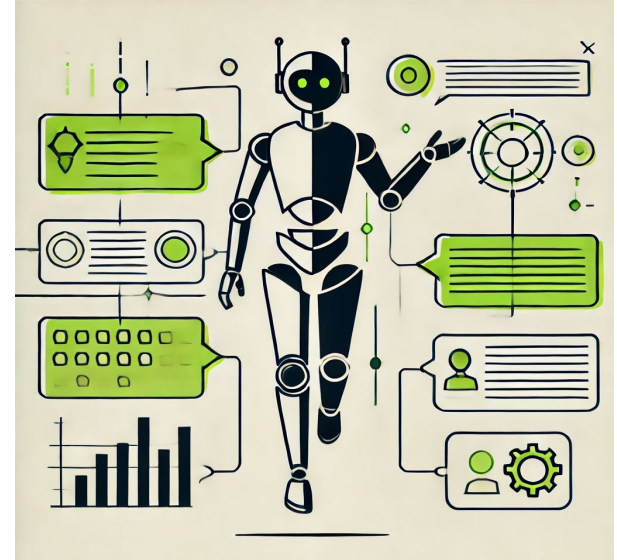
Quelle: DALL·E 2024-11-29 00.10.12 - A professional and graphic design featuring a chatbot figure receiving a set of instructions and contextual info

2. Kontext

Kontext	Beispiel	Pro-Tipp
Beschreibung (z.B. Fokus)	<i>„Erstelle fünf Prüfungsfragen für eine Einführungsvorlesung in Umweltwissenschaften mit Fokus auf Klimawandel. Nutze grundlegendes Wissen, ohne zu viele Fachbegriffe.“</i>	Chatbot bitten, weitere Infos abzufragen
Externe Daten (Datei, kopierte Inhalte)	<i>„Nutze das Dokument ‘Klimawandel_Grundlagen.pdf’, um fünf Prüfungsfragen zu erstellen. Halte dich ausschließlich an die dort enthaltenen Informationen.“</i>	Umfangreiche Dokumente aufteilen („Lost in the Middle“)
Beispiele (Few-Shot-Prompting)	<i>„Hier zwei Beispiel-Prüfungsfragen zum Thema Klimawandel: 1) ‘Welche Rolle spielt CO₂ für den Treibhauseffekt?’ 2) ‘Erkläre in drei Sätzen, warum die globale Erwärmung zu Extremwetterereignissen führt.’ Erstelle nun fünf ähnliche Fragen zum Thema Plastikverschmutzung.“</i>	je besser, je mehr Beispiele

3. Formale Vorgaben

- **Längenbeschränkung**
- **Format**, z.B. Stichpunkte, Tabelle
- **Nummerierung & Gliederung**, z.B. Einleitung, Hauptteil, Fazit
- **Sprache** / Stil definieren
- Keine überflüssigen Erklärungen



Quelle: DALL·E 2024-11-28 13.03.09 - A
simple and graphic representation of a
chatbot figure dividing tasks without any text.

Aktuelle Entwicklungen

Implementierung von Prompting-Methoden in KI-Modelle:

1. Automatische Prompt-Optimierung
2. Reasoning-Modelle (z.B: OpenAI o1, DeepSeek R1): *Chain of Thought, Tree of Thought* oder *Self-Consistency*
3. Effekt von *Role Prompting* in (neueren Modellen) strittig

⇒ Fokus stärker auf Auswahl des passenden **Modells** und **Datenbasis**

Praxisphase

Aufgabe - Einzelarbeit

Verbessern Sie den Prompt aus der ersten Aufgabe und vergleichen Sie anschließend die Ergebnisse. Nutzen Sie dafür z.B.:

- **Präzisere** und **eindeutigere** Formulierungen
- **Chained Prompting** (Aufteilung in Unteraufgaben) für umfangreiche Aufgaben wie Lehrvorbereitung, Projektplanung o.ä.
- Detaillierte **Kontextbeschreibung** (z.B. Hauptziel, Anforderungen, Einschränkungen)
- Nutzung von eigenen **externen Daten** oder hochgeladenen Dokumenten als Kontext
- **Few-Shot-Prompting** (Beispiele), wenn Sie bereits sehr gute ähnliche Texte haben, an denen sich der Chatbot orientieren soll

Sie haben dafür **10 Minuten** Zeit.



Quelle: DALL·E 2025-03-02
09.49.24 - A simple and clean
illustration of a teacher giving a
task to a chatbot.

Aufgabe - Gruppen

1. **Reflektieren** Sie die Ergebnisse mit eine*m Kolleg*in und geben Sie sich **Feedback** zur Verbesserung Ihrer Prompts.
2. Sammeln Sie weitere **Good Practices** im Prompting.

Sie haben dafür **10 Minuten** Zeit.



Quelle: DALL·E 2025-03-02
09.49.24 - A simple and clean
illustration of a teacher giving a
task to a chatbot.

Q&A

Quellen

Aufgaben und Kontext:

- Efrat, A., & Levy, O. (2020). The turking test: Can language models understand instructions?. arXiv preprint arXiv:2010.11982.
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12, 157-173.
- Mishra, S., Khashabi, D., Baral, C., Choi, Y., & Hajishirzi, H. (2021). Reframing Instructional Prompts to GPTk's Language. arXiv preprint arXiv:2109.07830.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint arXiv:2302.11382.

Beispiele (Few-shot Prompting)

- Logan IV, R., Balazevic, I., Wallace, E., Petroni, F., Singh, S., & Riedel, S. (2022). Cutting Down on Prompts and Parameters: Simple Few-Shot Learning with Language Models. Findings of the Association for Computational Linguistics: ACL 2022, 2824–2835 <https://doi.org/10.18653/v1/2022.findings-acl.222>
- Touvron, Hugo, et al. "Llama: Open and efficient foundation language models." arXiv preprint arXiv:2302.13971 (2023).
- Zhao, T. Z., Wallace, E., Feng, S., Klein, D., & Singh, S. (2021). Calibrate Before Use: Improving Few-Shot Performance of Language Models." arXiv. arXiv preprint arXiv:2102.09690.

Role Prompting:

- Zheng, M., Pei, J., Logeswaran, L., Lee, M., & Jurgens, D. (2024). When "A Helpful Assistant" Is Not Really Helpful: Personas in System Prompts Do Not Improve Performances of Large Language Models. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 15126-15154).
- Sander Schulhoff (2024) Revisiting Roles, URL: https://learnprompting.org/docs/intermediate/revisiting_roles (Zugriff: 04.03.2025)

Quellen

Modelle testen

- <https://www.buzzwoo.de/kuenstliche-intelligenz#test-arena> (Digitalagentur Buzzwoo)
- <https://lmarena.ai/> (Chatbot Arena is an open platform for crowdsourced AI benchmarking, hosted by researchers at UC Berkeley SkyLab and LMArena.)

Prompting Guides der Anbieter

- Gemini for Google Workspace: Prompting 101 (2024), URL: <https://services.google.com/fh/files/misc/gemini-for-google-workspace-prompting-guide-101.pdf> (Zugriff 02.03.2025)
- OpenAI Prompt engineering, URL: <https://platform.openai.com/docs/guides/prompt-engineering> (Zugriff: 02.03.2025)
- Anthropic Prompt Engineering, URL: <https://docs.anthropic.com/en/docs/build-with-claude/prompt-engineering/overview> (Zugriff: 02.03.2025)

Chatbots fürs wissenschaftliche Arbeiten: „Prompts – Leitfaden für die Kommunikation mit der generativen KI“ (S. 115-137)

- Bucher, Ulrich, Markus Schwarzer, and Kai Holzweißig. Künstliche Intelligenz und wissenschaftliches Arbeiten: ChatGPT & Co: Der Turbo für ein erfolgreiches Studium. Franz Vahlen, 2024.

RAG

- Ronny Frankenstein: Eigene Chatbots mit Retrieval Augmented Generation erstellen: https://www.heise.de/ratgeber/Eigene-Chatbots-mit-Retrieval-Augmented-Generation-erstellen-9748329.html?wt_mc=rss.red.ho.themen.k%C3%BCnstliche+intelligenz.beitrag_plus.beitrag_plus

Vielen Dank!

René Perfözl
rene.perfoelz@fu-berlin.de

Freie Universität
Lehren, Lernen und Prüfen mit
Künstlicher Intelligenz
ki@fu-berlin.de

AI Week: Organisatorisches
e-teaching@cedis.fu-berlin.de



AIQualify - Qualifizierung zu Künstlicher Intelligenz in der Hochschullehre

Koordination und Kontakt
Freie Universität Berlin
FUB-IT. Lehre, Studium, Forschung (ex-CeDiS)
E-Learning und E-Examinations (EEE)
Innestraße 24, D-14195 Berlin
e-teaching@cedis.fu-berlin.de
<https://www.cedis.fu-berlin.de/nel-ki>

VERANSTALTET VON:



IM RAHMEN EINES PROJEKTES VON:



GEFÖRDERT VON:

