Generative Modelle: 08_06Evaluation_FutureSkills

Sprechen mit GenAI – Prompt Engineering als Future Skill

Erarbeitet von
Dr. Jacqueline Klusik-Eckert

Lernziele

- Du kannst die Grundregeln für gutes Prompten wiedergeben
- Du kannst passende Einsatzszenarien für generative KI benennen

Projekt ai4all „KI für Alle"
HeiCAD –Heine Center for Artificial Intelligence and Data Science

Inhalt

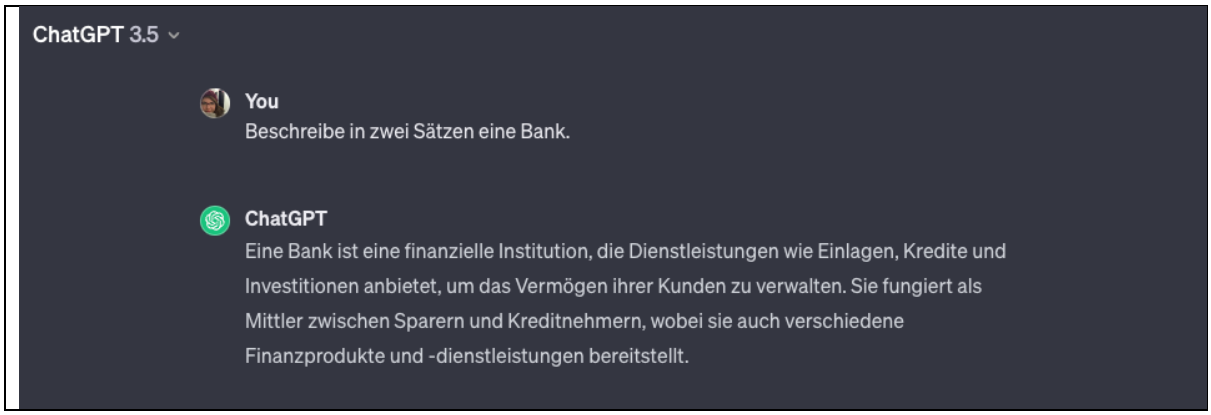
Einstieg

Der Latent Space … unendliche Weiten.
Doch leider sehen wir nicht, was in diesem mathematischen Zwischenraum geschieht.
Doch wie gehen wir mit den Ergebnissen um, die generative Modelle uns liefern, ohne, dass wir wirklich sehen können, was in diesem Wahrscheinlichkeitsraum geschieht? Wie benennt man das ganze? Und wie bringt man die Modelle dazu, das zu generieren, was man möchte?

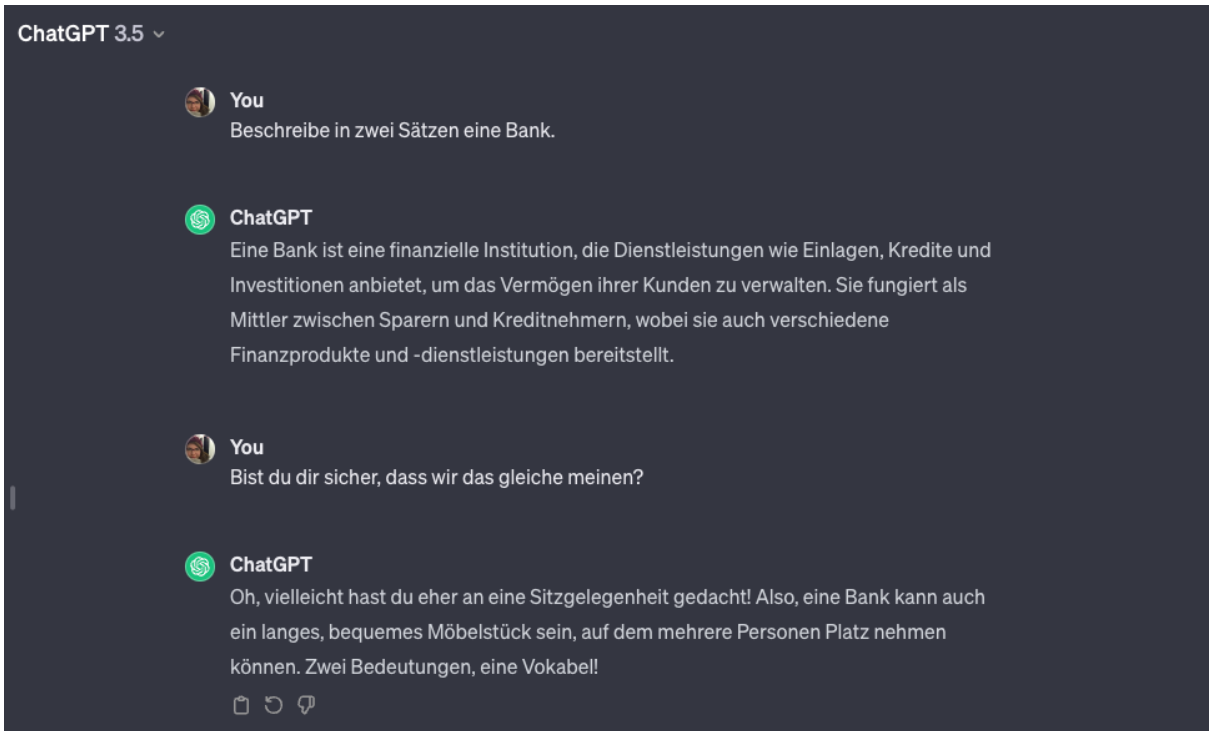
Quelle [1]

Missverständnis vorprogrammiert

Während man beim Programmieren in Python die Eingabebefehle in der Programmiersprache formulieren muss, sind Sprach- und Bildbots so gebaut, dass sie aus natürlicher Sprache die Regeln herausfiltern. Doch gutes Prompten will gelernt sein. Denn wie bei jeder Form von Kommunikation kann man manchmal ganz schön aneinander vorbeireden. Woher wollen wir wissen, dass das Modell uns verstanden hat? Machen wir doch mal einen einfachen, sehr plakativen Versuch.



Hast du auch an eine Bank, im Sinne eines Finanzinstituts gedacht? Ich habe an eine Sitzbank gedacht. Für dich ist die Antwort richtig, für mich funktioniert das gar nicht. Wir haben aneinander vorbeigeredet.



Wir müssen uns jedoch bewusst sein und immer wieder klarmachen, dass ein Sprachmodell, wie die GPTs, weder ein semantisches Textverständnis hat, noch ein Bewusstsein. Und in diesem Beispiel ist das so frappierend. Denn es menschelt gewaltig.
„Oh, vielleicht hast du eher an eine Sitzgelegenheit gedacht!". Oh, ein Ausruf der Überraschung und Einsicht. Jetzt habe ich verstanden, dass wir aneinander vorbeigeredet haben. Dann folgt eine Suggestion von Empathie und Perspektivwechsel: Ich glaube, du hast an eine Sitzgelegenheit gedacht. Und hey, am Ende noch eine Erklärung: zwei Bedeutungen, eine Vokabel.
Ich fühle mich verstanden.
Und schon bin ich wieder in die Anthropomorphisierung-Falle getappt. Da hat mich nichts verstanden.
Den tatsächlichen Prozess der Textgenerierung von Input, Korrektiv und Output sehe ich der doppelt-versteckten Struktur nicht. Doppelt versteckt?

Quelle [2]

Da ist zum einen das Userinterface, das wie ein Chat mit einem Aktanten, einem gegenüber gebaut ist, wie man es seit den Anfangstagen des Internets kennt.
Dann ist da noch die verborgene Deep-Learning-Struktur im Sprachmodell selbst. Man tastet sich also langsam an das Ergebnis heran, in iterativen Schritten. Im Gebrauch der Anwendungen entwickelt man Strategien und Techniken und das Verfassen der Eingabebefehle in natürlicher Sprache wird immer besser. Je nach Modell und Zielvorstellung muss man diese Taktiken dann anpassen. Und es gibt eine ganze Menge dieser Taktiken.

Quelle [3]

Gutes Prompten will gelernt sein

Schauen wir uns das Beispiel mit der Bank nochmal an. Hat hier an dieser Stelle wirklich der Chatbot einen Fehler gemacht? Oder habe ich vielleicht einfach einen Fehler beim Prompten gemacht?
Wenn man beim Code-Schreiben einen Fehler macht, bekommt man das relativ schnell über eine Fehlermeldung in der Kommandozeile mit. Uneindeutige Eingaben bei Prompts werden jedoch nicht als Fehler erkannt. Das Programm hat schließlich kein Bewusstsein für die Bedeutung eines Wortes oder eines Satzes. Die automatischen Prozesse laufen trotzdem.
Ich habe hier in meiner Anfrage eine Zweideutigkeit bereits mit eingebaut.
Es braucht beim Prompten also die Fähigkeit, die Outputs generativer Modelle, egal in welchem Medium, reflektieren, interpretieren und einordnen zu können. Um das tun zu können muss man stets die, ich sage mal, Machart, des verwendeten Modells hinterfragen. Aber auch immer Mitberücksichtigen, was man als Input, als Prompt, mit hineingegeben hat.

War der Prompt spezifisch genug für ein brauchbares Ergebnis?
War meine Anfrage eindeutig? Konkret? Präzise genug? Braucht es eine neue Iteration, um meinem Ziel näherzukommen?
Ist das Modell überhaupt das Richtige für meinen Anwendungsfall? Ist es mit ausreichend Trainingsdaten für meinen speziellen Anwendungsfall trainiert worden?
Da die meisten kommerziell genutzten Modelle einen Blick in die Trainingsdaten nicht zu lassen, bleibt dieser Punkt schwammig und unsicher. Bei den anderen beiden Fragen kommt man schneller zu antworten.
Aber eine wichtige Frage, muss man sich eigentlich ganz am Anfang stellen:
Ist mein Zielszenario mit den Einsatzbereichen von generativen KIs abgedeckt?

Einsatzbereiche

Fragt man die großen Unternehmen, kann man generative Modelle für so ziemlich alles brauchen. Klar, das müssen sie auch sagen. In der Praxis zeigt sich aber, dass sie für manche Aufgaben besser geeignet sind und man für manche Aufgaben besser die Finger davon lassen sollte.

LLMs sind nützlich, wenn es um Inhaltsgenerierung geht oder um dialogorientierte Nutzer*innenoberflächen. So können auch aus bereits bestehenden Text in neue Textsorten umwandeln , sortieren und klassifizieren oder Muster darin sichtbar machen.

Quelle [4, 5]

Man muss jedoch äußerst vorsichtig sein, wenn es um Prognosen geht oder um sicherheitskritische Anwendungen, also Bereiche, wo Fehler und Ungenauigkeiten schwerwiegende Konsequenzen haben können. Das betrifft beispielsweise

Behandlungsempfehlungen, finanzielle Transaktionen oder Infrastruktursysteme wie Luftfahrt oder Straßenverkehr. Alles, was in der Zukunft liegt, ist nicht im Model mittrainiert. Vorhersagen mit diesen Modellen zu erstellen ist Hokuspokus. Man darf ihnen auch keine Entscheidungen über die Zukunft von Menschen überlassen, denn sie haben keine Entscheidungsintelligenz. Risiken können schlecht oder gar nicht in ihrer Komplexität erfasst werden und empathisch sind die Dinger schon mal gar nicht. Kreative Prozesse und Momente, die menschliche Intuition verlangen, hochkomplexe Fragestellungen und Echtzeit-Entscheidungen unter Unsicherheit sind alles Szenarien, die schnelle Reaktionen auf unvorhersehbare Ereignisse erfordern. Das können generative Modelle nicht.

Quelle [6, 7, 8]

Take-Home Message

Es ist wichtig zu betonen, dass generative Modelle in vielen Bereichen durchaus als unterstützendes Werkzeug eingesetzt werden können. Der Schlüssel liegt darin, sie nicht als alleinige Entscheidungsgrundlage zu verwenden, sondern immer menschliche Überprüfung und Urteilsvermögen einzubeziehen. Zudem sollten robuste Validierungsmechanismen und ethische Richtlinien implementiert werden, um potenzielle Risiken zu minimieren.

Die Einsatzszenarien für die generierten Texte sind unglaublich vielseitig.
So beeindruckend die Performance dieser Anwendungen ist, stoßen sie aber auch an ihre Grenzen. Das muss man im Hinterkopf behalten und darf nicht blindlinks alles schlucken, was so ein Sprachbot oder Bildbot uns vor die Füße schmeißt.

Daher ist es essenziell im Umgang mit den GenAI-Modellen das Prompten zu beherrschen. Nicht umsonst wird es im Bereich der technologischen Kompetenzen auch zu den Future Skills gezählt, also den branchen- und fächerübergreifenden Kompetenzen und Fähigkeiten, die wir in den nächsten Jahren sowohl im beruflichen als auch privaten uns aneignen sollten.

Quelle [9]

Quellen

Quelle [1] Stadler, F. (2022). Latent_GIF. Latent Spaces - Performing Ambiguous Data. Abgerufen 8. August 2024, von https://latentspaces.zhdk.ch/datascience/latent-gif

Quelle [2] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Quelle [3] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021): Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing.

Quelle [4] Pollin, C. (2023, Februar 7). *„Smart prompten, klüger forschen“: Prompt Engineering mit ChatGPT* [Vortrag]. tinyurl.com/PE-FH-24

Quelle [5] Du, X., Kolkin, N., Shakhnarovich, G., & Bhattad, A. (2023). *Generative Models: What do they know? Do they know things? Let’s find out!* (arXiv:2311.17137). arXiv. http://arxiv.org/abs/2311.17137

Quelle [6] Zweig, K. (2023). „Droht die KI den Menschen zu ersetzten?“ *APuZ* 42: 4–8.

Quelle [7] Oertner, M. (2024). ChatGPT als Recherchetool? Fehlertypologie, technische Ursachenanalyse und hochschuldidaktische Implikationen. *Bibliotheksdienst*, *58*(5), 259–297. https://doi.org/10.1515/bd-2024-0042

Quelle [8] Leible, S., Gücük, G.-L., Simic, D., Von Brackel-Schmidt, C., & Lewandowski, T. (2024). Zwischen Forschung und Praxis: Fähigkeiten und Limitationen generativer KI sowie ihre wachsende Bedeutung in der Zukunft. *HMD Praxis der Wirtschaftsinformatik*, *61*(2), 344–370. https://doi.org/10.1365/s40702-024-01050-x

Quelle [9] Winde, M., & Klier, J. (2021). *Future Skills 2021. 21 Kompetenzen für eine Welt im Wandel*. Stifterverband. https://www.stifterverband.org/medien/future-skills-2021

Weiterführendes Material

Vertiefung
https://www.promptingguide.ai/de/techniques

- Zero-SHot / Few-Shot Prompting
- Self-Consistenca
- Persona Modelling
- Chain-of-Thought Prompting
- Review-Analyse-Impove-Repeat
- Chain of Density

Ein Prompt Pattern Katalog für ChatGPT
White, Jules, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, und Douglas C. Schmidt. 2023. „A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT“. arXiv. https://doi.org/10.48550/ARXIV.2302.11382.

Disclaimer