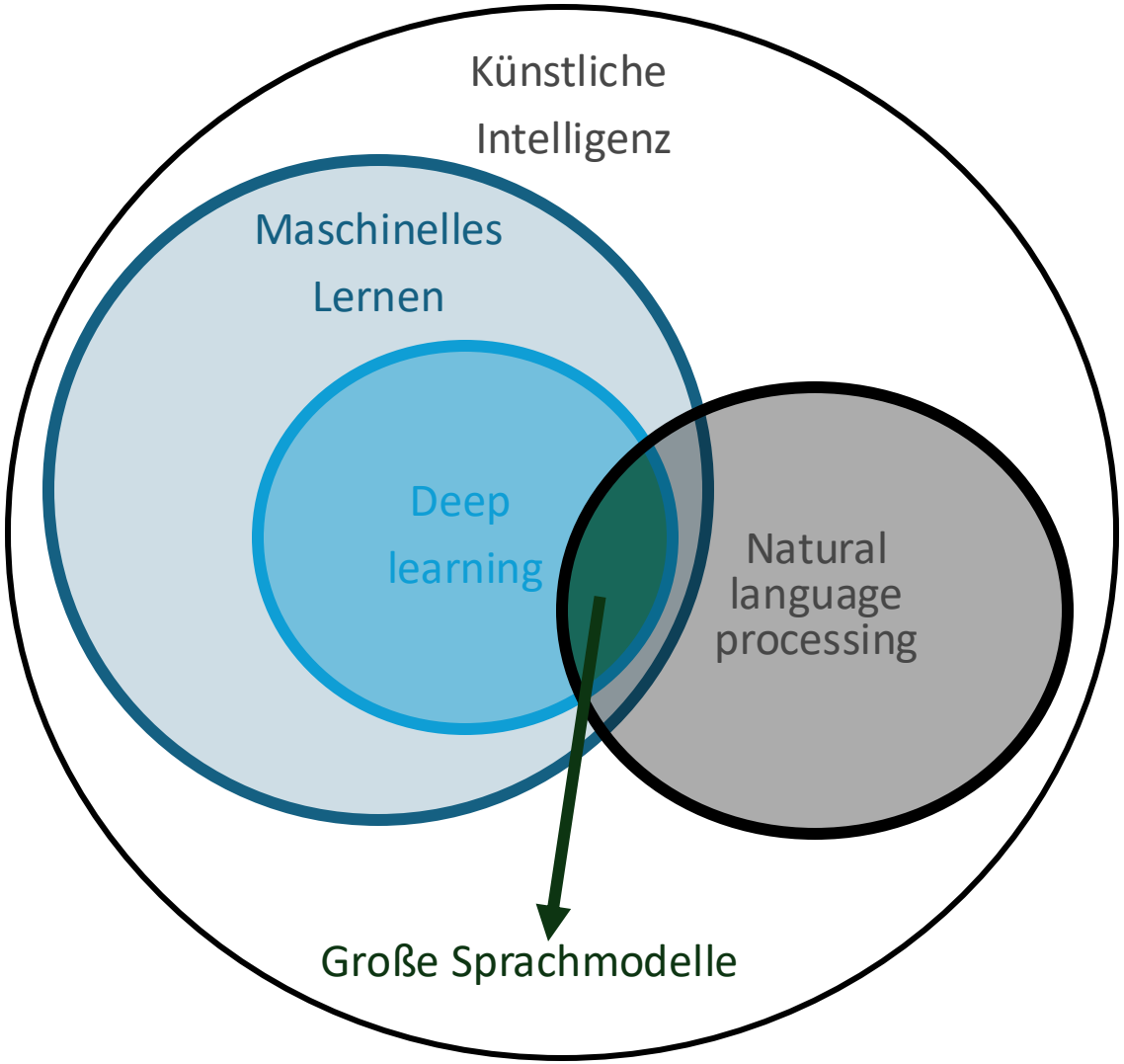
Maschinelles Lernen

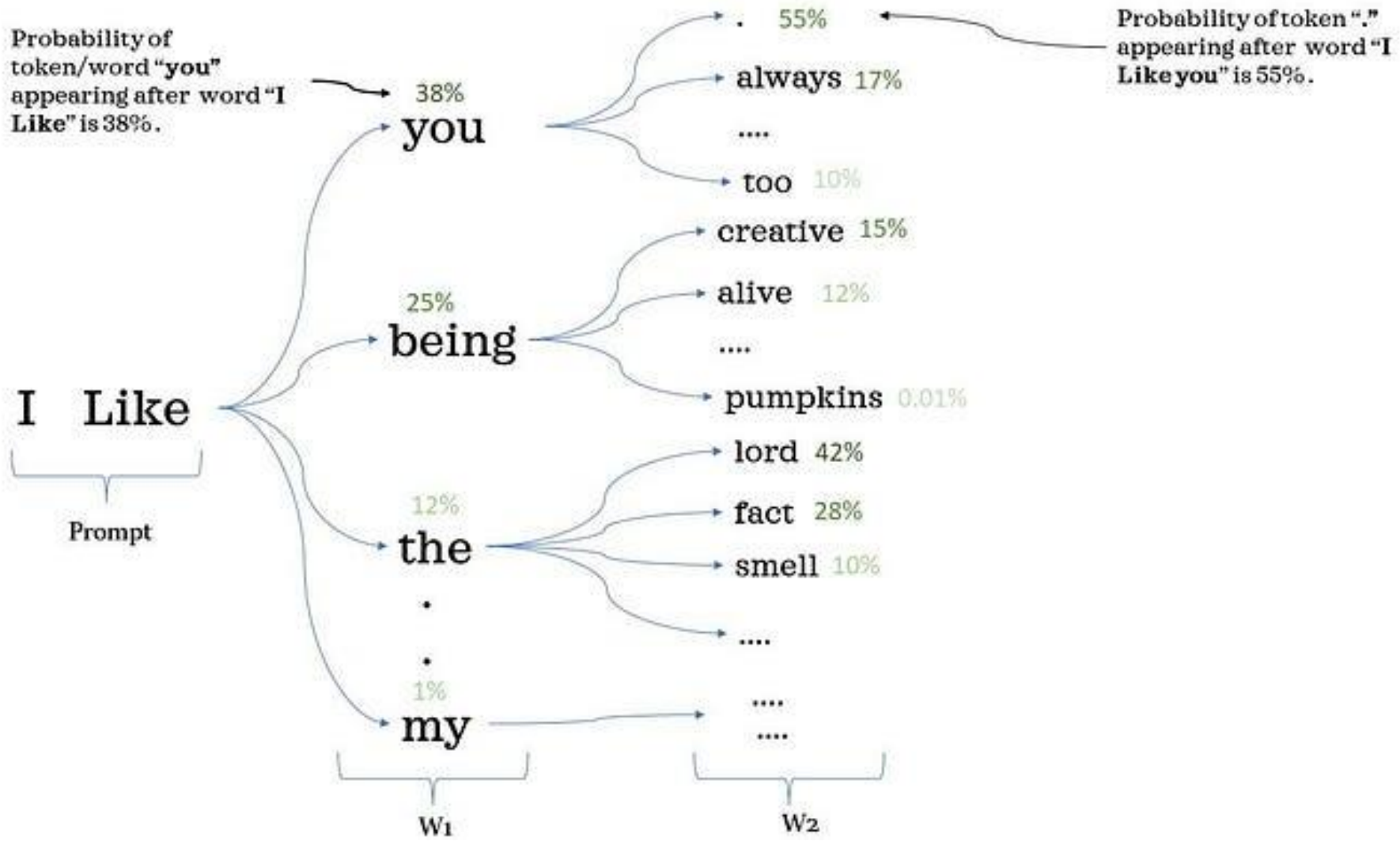
Vorlesung 3

Grundlagen großer Sprachmodelle



- LLMs sind KI-Systeme, die anhand riesiger Mengen von Texten trainiert werden, um Text zu verstehen und zu generieren.
- Sie lernen aus den Daten **Muster, Grammatik** und **Semantik**.
- LLMs sind **mathematische Algorithmen**, die auf Wahrscheinlichkeiten basieren. Sie verfügen über keinerlei Wissen.
- LLMs sind **sequenzielle** Vorhersagemodelle. Sie **sagen das nächste** Wort oder Token in einer Sequenz **voraus,** basierend auf dem Kontext der vorherigen Wörter oder Token in der Sequenz.

Das nächste Token wird basierend auf seiner Wahrscheinlichkeit generiert!



Quelle: https://masteringllm.medium.com/demystifying-the-temperature-parameter-a-visual-guide-to-understanding-its-role-in-large-language-d9e8ea4b9956

Beispiele

Was man bei der Auswahl eines Modells beachten sollte:

- Preise
- Token-Limit
- Trainings-Cutoff-Datum (auch für Modelle mit Webbrowsing-Fähigkeiten)
- Fähigkeit, Bilder zu verstehen

- Versuche, „Störgeräusche” zu ignorieren oder zu überlesen
- Modelle ändern sich mit jeder Version
- Doomers und Techbros sind immer am lautesten

Gemini Ultra: Benchmark-Leistung

Gemini: A Family of Highly Capable Multimodal Models

	Gemini Ultra	Gemini Pro	GPT-4	GPT-3.5	PaLM 2-L	Claude 2	Inflect-ion-2	Grok 1	LLAMA-2
MMLU Multiple-choice questions in 57 subjects (professional & academic) (Hendrycks et al., 2021a)	**90.04%** CoT@32*	79.13% CoT@8*	87.29% CoT@32 (via API**)	70% 5-shot	78.4% 5-shot	78.5% 5-shot CoT	79.6% 5-shot	73.0% 5-shot	68.0%***
	83.7% 5-shot	71.8% 5-shot	86.4% 5-shot (reported)						
GSM8K Grade-school math (Cobbe et al., 2021)	**94.4%** Maj1@32	86.5% Maj1@32	92.0% SFT & 5-shot CoT	57.1% 5-shot	80.0% 5-shot	88.0% 0-shot	81.4% 8-shot	62.9% 8-shot	56.8% 5-shot
MATH Math problems across 5 difficulty levels & 7 subdisciplines (Hendrycks et al., 2021b)	**53.2%** 4-shot	32.6% 4-shot	52.9% 4-shot (via API**)	34.1% 4-shot (via API**)	34.4% 4-shot	—	34.8%	23.9% 4-shot	13.5% 4-shot
			50.3% (Zheng et al., 2023)						
BIG-Bench-Hard Subset of hard BIG-bench tasks written as CoT problems (Srivastava et al., 2022)	**83.6%** 3-shot	75.0% 3-shot	83.1% 3-shot (via API**)	66.6% 3-shot (via API**)	77.7% 3-shot	—	—	—	51.2% 3-shot
HumanEval Python coding tasks (Chen et al., 2021)	**74.4%** 0-shot (IT)	67.7% 0-shot (IT)	67.0% 0-shot (reported)	48.1% 0-shot	—	70.0% 0-shot	44.5% 0-shot	63.2% 0-shot	29.9% 0-shot
Natural2Code Python code generation. (New held-out set with no leakage on web)	**74.9%** 0-shot	69.6% 0-shot	73.9% 0-shot (via API**)	62.3% 0-shot (via API**)	—	—	—	—	—
DROP Reading comprehension & arithmetic. (metric: F1-score) (Dua et al., 2019)	**82.4** Variable shots	74.1 Variable shots	80.9 3-shot (reported)	64.1 3-shot	82.0 Variable shots	—	—	—	—
HellaSwag (validation set) Common-sense multiple choice questions (Zellers et al., 2019)	87.8% 10-shot	84.7% 10-shot	**95.3%** 10-shot (reported)	85.5% 10-shot	86.8% 10-shot	—	89.0% 10-shot	—	80.0%***
WMT23 Machine translation (metric: BLEURT) (Tom et al., 2023)	**74.4** 1-shot (IT)	71.7 1-shot	73.8 1-shot (via API**)	—	72.7 1-shot	—	—	—	—

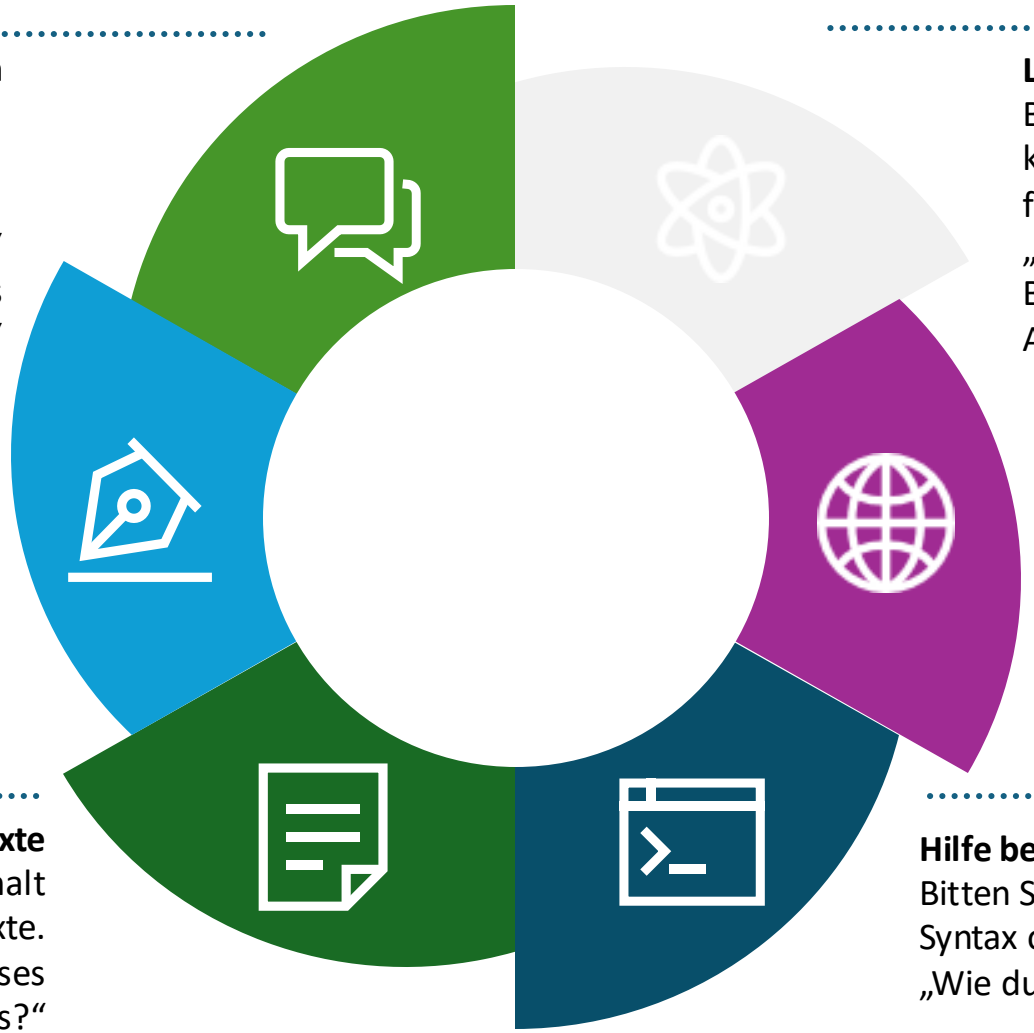
Table 2 | Gemini performance on text benchmarks with external comparisons and PaLM 2-L.

Wie kann mir GenAI helfen? Was kann GenAI tun?

Fragen und Antworten
Fragen nach Definitionen, Erklärungen, Fakten, Antworten oder Empfehlungen zu bestimmten Themen.
„Was ist der Treibhauseffekt?“
„Welche Strategien gibt es für das Zeitmanagement?“

Kreatives Schreiben
Anregungen für kreatives Schreiben anfordern, Ideen für Geschichten oder fantasievolle Inhalte.
„Schlage mir eine einzigartige Idee für ein neues Produkt vor.“

Zusammenfassen langer Texte
Fragen nach einer Übersicht über den Inhalt sehr umfangreicher Texte.
„Was sind die wichtigsten Punkte dieses Dokuments?“



Lernen und Bildung
Bitten Sie um Erklärungen zu komplexen Konzepten oder um Schritt-für-Schritt-Anleitungen.
„Führe mich durch den Prozess der Berechnung der Algorithmenkomplexität.“

Sprachübersetzung
Übersetzen Sie Ausdrücke oder Sätze von einer Sprache in eine andere.

Hilfe beim Programmieren
Bitten Sie um Hilfe bei Programmierproblemen, Syntax oder Programmierkonzepten.
„Wie durchlaufe ich eine Liste in Python?“

Definition und Bedeutung

Was ist ein Prompt?

- Prompts sind Anweisungen, die einem LLM gegeben werden, um dessen Verhalten zu steuern und relevante Ergebnisse zu generieren.
- Beispiel: „Was sind die Ursachen der globalen Erwärmung?“

Was ist Prompt Engineering?

- Leitet Modelle an, um gewünschte Ergebnisse effektiv zu generieren.
- Anpassen der Modellausgaben durch maßgeschneiderte Eingabeaufforderungen an die erwünschten Ziele.
- Benutzerdefinierte Prompts liefern den erforderlichen Kontext und die notwendigen Einschränkungen mit minimalem Token-Aufwand.
- Gewährleistet genaue, relevante und kohärente Antworten von KI-Modellen.

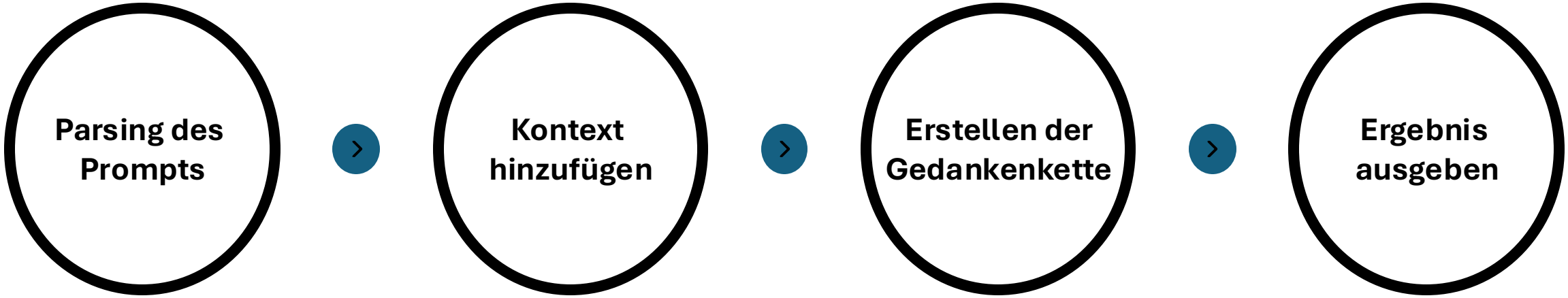
Die Bedeutung gut gestalteter Prompts

- **Guter Prompt**: „Fasse die Hauptursachen des Klimawandels und deren Auswirkungen auf die Umwelt zusammen. Zitiere relevante, seriöse Daten und Quellen.“
- **Schlechter Prompt**: „Erzähle mir etwas über den Klimawandel.“

Model reasoning

reasoning (noun): the process of thinking about things in a logical way; opinions and ideas that are based on logical thinking

(Oxford English Dictionary)



In die Sprache des logisch-deduktiven Denkens umwandeln

Daten aus dem eigenen Speicher, einer Internetsuche oder einem vorgegebenen Kontext

Anwendung von Schlussfolgerungsregeln: „Wir wissen A und B, daher C“

Anwendung der üblichen Textgenerierung

Eigentlich…

Was genAI-Tools tun, ist nicht wirklich Nachdenken / reasoning.

Richtlinien für das Verfassen wirkungsvoller Prompts

Sei klar und spezifisch

Vermittle klar die gewünschte Aufgabe oder Information, die das Modell generieren soll.

Kontextinformationen

Stelle dem Sprachmodell ausreichend Kontext zur Verfügung, damit es sinnvolle Antworten generieren kann.

Ton, Format, Rolle und Einschränkungen

Gebe den gewünschten Ton an (z. B. locker, formell, usw).

Definiere ein Format (z. B. Tabelle, Stichpunkte, Aufsatz).

Gebe eine Rolle / Perspektive an (Experte, Laie, Kind, ...).

Gebe eine Wort- oder Zeichenanzahl an

Halte dich an die Modellbeschränkungen

Beachte die Grenzen und Fähigkeiten des Sprachmodells.

Zitate und Statistiken

Fordere die Angabe der Quelle für Informationen

Fordere Statistiken oder Daten als Beweis von Behauptungen an

Bitte darum, bestimmte Quellen aufzunehmen bzw auszulassen.

Experimentiere und iteriere

Teste verschiedene Prompts und Parameter, um die für die vorgegebene Aufgabe effektivsten zu finden.

Behandeln von Biases

Sei dir potenzieller Biases bewusst und entwerfe Prompts, die Fairness und Inklusivität fördern.

Bewerte die Leistung der Prompts

Bewerte die Leistung der Prompts kontinuierlich und passe sie bei Bedarf an.

Zusammenfassung, Paraphrasierung, Stil, Ton

Anweisungen für Zusammenfassung

Quelle angeben
„Fasse den **folgenden Artikel** über *[Thema]* in einem kurzen Absatz zusammen.“

Wichtige Punkte
„Erstelle eine Zusammenfassung mit Stichpunkten, ***die die wichtigsten Punkte*** aus dem Text zu [Thema] ***hervorheben***.“

Längenbeschränkungen
„Fasse dieses Dokument in einer ***3-Satz-Zusammenfassung*** zusammen.“

Fokus auf das Ziel
„Erstelle eine kurze Zusammenfassung, die den ***Kern*** des [Berichts/Dokuments] zum Thema [Thema] erfasst.“

Relevanz hervorheben
„Fasse ***die relevanten Abschnitte*** der [Quelle] zusammen, die sich auf [Aspekt] beziehen.“

Anweisungen zum Paraphrasieren

Ursprünglicher Sinn
„Schreibe den folgenden Absatz unter Beibehaltung der ursprünglichen Bedeutung um.“

Änderung des Stils
„Formuliere diesen Text in einem formelleren / informelleren Ton um.“

Thematisches Umschreiben
„Drücke die Ideen in diesem Absatz mit anderen Worten aus.“

Synonyme Sprache
„Gebe eine alternative Formulierung für die hervorgehobenen Sätze an.“

Beziehe dich auf eine bestimmte Wissensdatenbank
„Verwende nur Daten, die in [Quellen] vorhanden sind. Verwende keinerlei andere Informationsquellen.“

Kontrolle von Stil und Ton im generierten Text

Stilanweisungen
Gebe explizite Anweisungen, um den Stil zu steuern.

Beschreibende Formulierungen
Verwende Adjektive, die den gewünschten Ton widerspiegeln, wie z. B. warm, professionell oder leger

Beispielsätze
Füge Beispiele im gewünschten Stil hinzu, um dem Modell zu zeigen, was man erwartet.

Vergleiche
Verwende Gleichnisse und Metaphern, die dem gewünschten Stil entsprechen, um das Modell anzuleiten.

Feedback-Schleife
Wenn die anfängliche Ausgabe nicht dem gewünschten Stil entspricht, gebe Feedback, wiederhole die Anweisungen und/oder formuliere sie um.

Beispiel: Was ist die Ursache des Klimawandels?

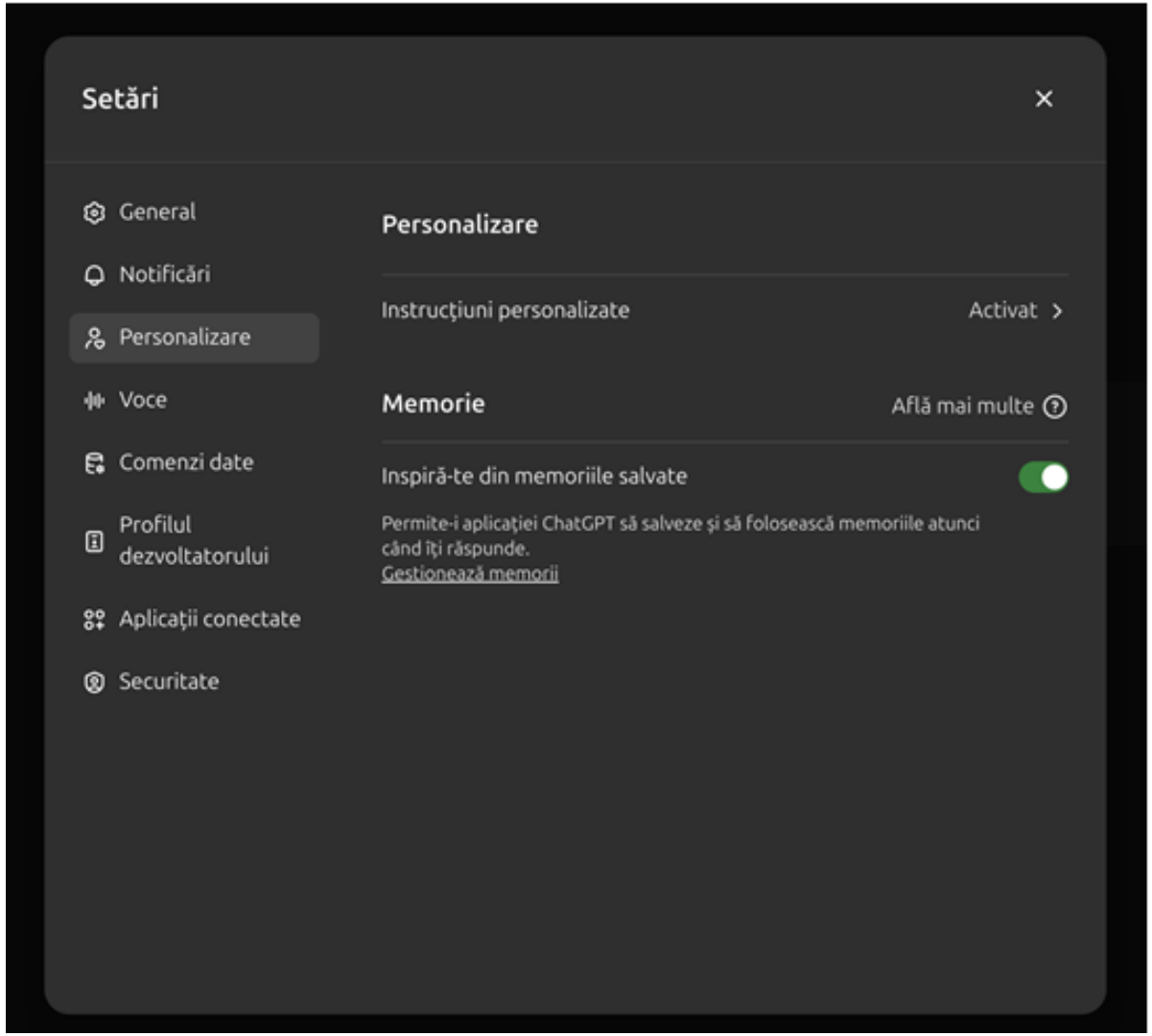
- Sei klar und spezifisch
  - Was sind die ***Hauptursachen*** des Klimawandels und ***seine Auswirkungen in Deutschland***?
- Hintergrundinformationen
  - ***Im Jahr 2021 kam es in Deutschland zu dramatischen Überschwemmungen*.** Was sind die Hauptursachen des Klimawandels und seine Auswirkungen in Deutschland?
- Ton, Format, Rolle und Einschränkungen
  - Im Jahr 2021 kam es in Deutschland zu dramatischen Überschwemmungen. Was sind die Hauptursachen des Klimawandels und seine Auswirkungen in Deutschland? ***Behalte einen objektiven und sachlichen Ton, verwende Stichpunkte, schreibe für einen Experten und beschränke dich auf maximal 100 Wörter*.**
- Zitate und Statistiken
  - Im Jahr 2021 gab es Überschwemmungen in Deutschland. Was sind die Hauptursachen des Klimawandels und seine Auswirkungen in Deutschland? Behalte einen objektiven und sachlichen Ton, verwende Stichpunkte, schreibe für einen Experten und beschränke dich auf maximal 100 Wörter. ***Füge Zitate und Statistiken hinzu, um die Behauptungen zu beweisen*.**
- Biases vermeiden
  - Im Jahr 2021 gab es Überschwemmungen in Deutschland. Was sind die Hauptursachen des Klimawandels und seine Auswirkungen in Deutschland? Behalte einen objektiven und sachlichen Ton, verwende Stichpunkte, schreibe für einen Experten und beschränke dich auf maximal 100 Wörter. Füge Zitate und Statistiken hinzu, um die Behauptungen zu beweisen. ***Verwende eine inklusive Sprache, die unterschiedliche Standpunkte und Erfahrungen berücksichtigt und Unsicherheiten anerkennt*.**

Permanenter Modellspeicher

Zum Beispiel…

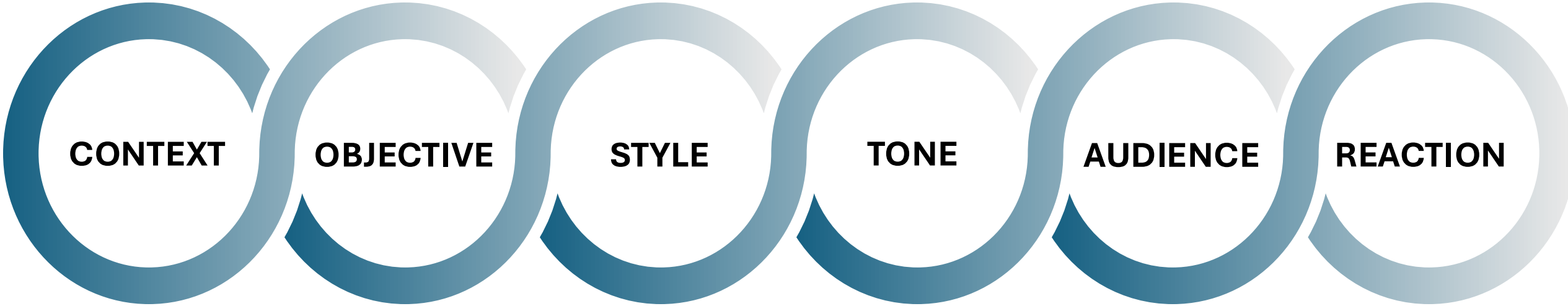
- „Benutzer möchte Schweizerdeutsch-Stil”
- „Benutzer möchte ausschließlich metrische Einheiten”
- „Der Hintergrund des Benutzers ist […]”

Gespeicherte Speicher können manuell verwaltet werden



Verwende bei Bedarf die Funktion „vorläufige Diskussion”.

Zusammenfassung:
Das COSTAR-Framework



Mögliche Strategien

N-Shot-Prompting

Erstelle von Anfang an ein Prompt mit allen notwendigen Informationen. Gebe Beispiele (falls notwendig) und kontrolliere Stil, Kontext, Tonfall, Rolle usw.

Chain-of-thought Prompting

Leite die KI Schritt für Schritt. Stelle die Informationen, die das Modell benötigt, in aufeinanderfolgenden Schritten bereit, d. h. gebe wiederholte Eingabeaufforderungen. Simuliere einen Dialog, den du in die gewünschte Richtung lenkst.

Funktioniert am besten bei großen und komplexen Aufgaben. Probiere es beim Brainstorming von Ideen aus!

Risiken und Einschränkungen

Biases und Ungenauigkeitenz

Ungenaue Informationen aufgrund von Trainingsdaten, die möglicherweise gesellschaftliche Biases widerspiegeln oder falsche Antworten liefern.

Mangelndes Kontextverständnis

Es kann sein, dass der Kontext in längeren Gesprächen nicht vollständig verstanden wird, was zu irrelevanten oder unsinnigen Antworten führt.

Halluzinationen und Fehlinformationen

Generiert Informationen, die plausibel klingen, aber völlig falsch sind, sodass eine Überprüfung der Fakten unerlässlich ist.

Abhängigkeit von Prompts

Modelle sind stark auf gut formulierte Prompts angewiesen, um die gewünschten Ergebnisse zu erzielen.

Ressourcenintensiv

Das Training und die Verwendung großer Modelle erfordern erhebliche Rechenleistung und Energie.
Training: so viel Strom wie eine Kleinstadt!
Eine Interaktion: so viel wie 1 Stunde Glühbirnenverbrauch!

Datenschutz und Sicherheit

Die Nutzer wissen nicht wirklich, wer auf ihre Daten zugreifen und sie einsehen kann. Wo (wenn überhaupt) werden sie gespeichert und verarbeitet? Was passiert mit sensiblen Informationen? Wo befinden sich das Modell und der Server physisch?

Weitere Tipps

Qualität der Informationen

Sage ihm, es solle relevante klärende Fragen stellen, anstatt etwas zu erfinden oder Lücken zu füllen.

Dies ist umso wichtiger, wenn man in einem sehr spezifischen Kontext arbeitet!

GenAI ist keine Online-Suche

Das Ziel von LLMs ist nicht, Fakten zu recherchieren. Sie können eine Internetsuche nicht ersetzen. Verwende LLMs, um Informationen zu bearbeiten, zusammenzufassen, übersetzen oder verdeutlichen, nicht um Informationen anzubieten.

Dialogverschmutzung

Wechsele während eines Dialogs nicht das Thema.

Beginne einen neuen Dialog, da man sonst den Kontext und das Gedächtnis der KI in Bezug auf das aktuelle Thema beeinträchtigt.

Business Standard, 9. Oktober 2025

Business Standard

HOME | TECHNOLOGY | TECH NEWS | TECH REVIEWS | GADGETS | E-PAPER | DECODED | OPINIC

Home / Technology / Tech News / Deloitte's AI fiasco: Why chatbots hallucinate and who else got caught

Deloitte's AI fiasco: Why chatbots hallucinate and who else got caught

As Deloitte returns part of its fee for a flawed AI-assisted government report in Australia, the case spotlights how generative AI fabricates facts and how it's not the first time this has happened

Über diese Geschichte wurde in zahlreichen Mainstream-Medien berichtet.

Auswertung der Prompts

- Bewusstsein für die Grenzen des Modells
  - LLMs sind probabilistische Algorithmen, die eine Fehlerquote aufweisen.
  - Sie haben nicht immer Zugriff auf die aktuellsten Informationen.
  - Sie sind nicht die absolute Wahrheit, und ihre Ausgabe kann durch die Eingabe manipuliert werden.
- Genauigkeit und sachliche Richtigkeit
  - Überprüfe die Genauigkeit der generierten Informationen anhand zuverlässiger Quellen. Überprüfe sie auf sachliche Fehler oder irreführende Inhalte.
- Relevanz für die Aufgabe
  - Bewerte, wie gut der generierte Inhalt die beabsichtigte Aufgabe oder Frage erfüllt.
  - Beurteile, ob die bereitgestellten Informationen mit dem erwünschten Ziel übereinstimmen.
- Bewertung der Biases
  - Untersuchen Sie Eingabeaufforderungen und Antworten auf mögliche versteckte Biases.
  - Identifiziere Inhalte, die irreführende Informationen perpetuieren könnten.
  - Beachte unterschiedliche Perspektiven.

Gruppenübung

1. Erstelle eine Präsentation über die Geschichte der künstlichen Intelligenz.
   - **Möglicher Beispielprompt für den Einstieg:** „Erstelle eine Powerpoint-Präsentation zum Thema [Thema] mit [n] Folien. Verfasse sie ausschließlich auf der Grundlage deines eigenen Wissens, ohne Platzhalter und ohne Online-Recherche.“
   - Experimentiert mit eurem Prompt, fügt zusätzlichen Kontext und/oder Anforderungen hinzu und versucht, auch Hinweise für das Design hinzuzufügen (Titelfolie, Textumfang, Zielgruppe usw.)
   - Wiederholt dies mehrmals
2. Erstelle **eine Strategie** für **die digitale Transformation** eines Möbelherstellers.
3. Stelle dem Bot eine **Nischenfrage** aus einem Bereich, in dem du dich sehr gut auskennst. Analysiere die Antwort kritisch und teile deine Erkenntnisse mit der Gruppe.

Stellt eure Ergebnisse den Anderen vor: was war gut, schlecht, lustig oder nervig?

Schlussfolgerungen und Ausblick

- Genauso wie Suchmaschinenoptimierung vor Google einfach kein Ding war, wird Prompt Engineering jetzt zu einem Thema welches für alle relevant ist
- GenAI ist ein großartiges Werkzeug – setze es richtig ein! Und verwende es nur wenn notwendig. „Wenn man nur einen Hammer hat, sieht man überall Nägel". Braucht dein Anwendungsfall (oder dein Kunde) es wirklich, oder folgest du nur dem Hype?