Chain-of-Thought-Prompting

Wann „Schritt für Schritt“ noch hilft, und wann der Denk-Modus das übernimmt

Chain-of-Thought-Prompting (CoT) lässt ein Sprachmodell seinen Lösungsweg ausschreiben, bevor es antwortet. Ausgelöst hat das 2022 der Satz **„Let's think step by step“.** Seit ChatGPT, Claude und Gemini als Reasoning-Modelle einen eigenen **Denk-Modus** haben, der diesen Prozess automatisch übernimmt, wirkt die klassische Anweisung nicht mehr überall gleich.

WOHER DIE TECHNIK KOMMT

FEW-SHOT-COT
Vorgerechnete Beispielaufgaben zeigen dem Modell den Lösungsweg, bevor es die eigentliche Aufgabe bearbeitet.
Wei u. a., Google, 2022

ZERO-SHOT-COT
Ein einziger Satz reicht, ganz ohne Beispiele.
„Let's think step by step“, Kojima u. a., 2022

Gemessen 2022 an InstructGPT und PaLM: MultiArith-Test 17,7 auf 78,7 Prozent richtig, GSM8K-Mathe-Test 10,4 auf 40,7 Prozent.

WAS DIE WHARTON-STUDIE 2025 GEMESSEN HAT

Modelltyp	Punktgewinn durch die Anweisung
Ohne Denk-Modus	**7,1 bis 13,5** Punkte mehr, Ausnahme GPT-4o-mini +4,4, nicht signifikant
Mit Denk-Modus	+2,9 und +3,1 Punkte, beim dritten Modell **−3,3** Punkte

Meincke u. a., Wharton School: 8 Modelle, 198 Naturwissenschafts-Fragen, je 25-mal gestellt. Reasoning-Antworten dauerten mit der Anweisung 20 bis 80 Prozent länger.

OHNE DENK-MODUS

- **Bleibt nützlich,** besonders bei Rechenaufgaben und mehrstufiger Logik

DER ENTSCHEIDENDE SATZ
„Zeige jeden Rechenschritt einzeln, bevor du den Endbetrag nennst.“
Jeder Zwischenwert steht offen da.

MIT DENK-MODUS

- **Ziel und Prüfkriterien beschreiben,** statt eine feste Schrittliste vorzugeben

DER ENTSCHEIDENDE SATZ
„Begründe deine Einschätzung nachvollziehbar und schlage eine konkrete nächste Handlung vor.“
Den Weg dorthin überlässt du dem Modell.

ZEIGEN DENKSCHRITTE, WAS DIE KI WIRKLICH DENKT?

Anthropic 2025: Ein Modell erwähnte einen genutzten Hinweis im Schnitt nur in 25 Prozent der Fälle, bei ausgenutzten Lücken unter 2 Prozent. **Sichtbare Denkschritte beweisen also nicht, wie die KI zur Antwort kam.**

HÄUFIGE FEHLANNAHMEN

„Schritt für Schritt hilft immer“
Bei Reasoning-Modellen ist er minimal oder negativ, kostet zusätzliche Zeit.

„Sichtbare Denkschritte erklären die Entscheidung“
Sie spiegeln nicht immer die tatsächlichen Einflussfaktoren wider.

„Ohne explizite Anweisung denkt das Modell nicht nach“
Reasoning-Modelle denken intern, unabhängig von einer Aufforderung.

„Mehr Anweisungen bringen bessere Ergebnisse“
Klarheit von Ziel und Prüfkriterien zählt mehr als die Schrittfolge.

MERKSÄTZE

Ohne Denk-Modus hilft Schritt für Schritt noch spürbar.

Mit Denk-Modus zählen Ziel und Prüfkriterien mehr.

Denkschritte beweisen nicht, wie die KI antwortet.