

Sprachmodelle richtig nutzen

Mit einem Sprachmodell (LLM) interagiert man, indem man ein Stück Text als Handlungsaufforderung (prompt) überliefert. Dieser Text wird dann auf Basis von statistischen Vorhersagen vervollständigt.

Mit einfachen Aufforderungen kann man viel erreichen, aber die Qualität der Ergebnisse hängt stark davon ab, wie viele Informationen man bereitstellt und wie gut die Aufforderung formuliert ist.

Eine Aufforderung sollte Informationen wie die Anweisung oder Frage, die man an das Modell stellt, enthalten. Darüber hinaus sind weitere Details wie Kontext oder Beispiele hilfreich. Man kann diese Elemente nutzen, um das Sprachmodell effektiver anzuleiten und so die Qualität der Ergebnisse zu verbessern.

Hier ein paar Tipps, wie man Sprachmodelle effektiv nutzt. Diese Anleitung ist natürlich nicht erschöpfend sondern soll ein erstes Gefühl für den generellen Umgang mit LLMs vermitteln.

Strategien für bessere Ergebnisse

Klare Anweisungen geben

Diese Modelle können Ihre Gedanken nicht lesen. Wenn die Ausgaben zu lang sind, bitten Sie um kurze Antworten. Wenn die Ausgaben zu einfach sind, bitten Sie um Texte auf Expertenniveau. Wenn Ihnen das Format nicht gefällt, nennen Sie das gewünschte Format. Je weniger das Modell raten muss, was man möchte, desto wahrscheinlicher wird man es erhalten.

Taktiken:

- Details und Kontext in die Anfrage aufnehmen, um relevantere Antworten zu erhalten
- Das Modell bitten, eine Rolle zu übernehmen ("Antworte als Experte für ...")
- Trennzeichen verwenden, um verschiedene Teile der Eingabe deutlich zu machen
- Die Schritte angeben, die zur Vollendung einer Aufgabe erforderlich sind
- Beispiele geben
- Die gewünschte Länge der Ausgabe spezifizieren
- ...

Referenzen und Beispiele bereitstellen

Sprachmodelle können fälschlicherweise Antworten erfinden, beispielsweise wenn sie nach esoterischen Themen oder nach Zitaten und URLs gefragt werden. Ähnlich wie ein Spickzettel einem Schüler bei einer Prüfung helfen kann, kann das Bereitstellen von Referenztexten diesen Modellen helfen, mit weniger Erfindungen zu antworten.

Taktiken:

- Das Modell anweisen, ähnlich einem Referenztext zu antworten
- Das Modell anweisen, mit Zitaten aus einem Referenztext zu antworten
- ...

Komplexe Aufgaben in einfachere Teilaufgaben unterteilen

So wie es in der Softwareentwicklung gute Praxis ist, ein komplexes System in eine Reihe von modularen Komponenten zu zerlegen, gilt dies auch für Aufgaben, die an ein Sprachmodell gestellt werden. Komplexe Aufgaben haben tendenziell höhere Fehlerquoten als einfache Aufgaben. Darüber hinaus können komplexe Aufgaben oft als Arbeitsablauf einfacherer Aufgaben neu definiert werden, wobei die Ausgaben früherer Aufgaben verwendet werden, um die Eingaben für spätere Aufgaben zu konstruieren.

Taktiken:

- Intent-Klassifikation verwenden, um die relevantesten Anweisungen für eine Benutzeranfrage zu identifizieren
- Für Dialoganwendungen, die sehr lange Gespräche erfordern, vorherige Dialoge zusammenfassen oder filtern
- Lange Dokumente stückweise zusammenfassen und eine vollständige Zusammenfassung rekursiv erstellen
- ...

Dem Modell Zeit zum "Nachdenken" geben

Wenn man gefragt wird, 17 mit 28 zu multiplizieren, weiß man dies vielleicht nicht sofort, kann es aber mit etwas Zeit herausfinden. Ähnlich machen Modelle mehr Denkfehler, wenn sie versuchen, sofort zu antworten, anstatt sich Zeit zu nehmen, um eine Antwort zu erarbeiten. Das Erfragen einer "Gedankenkette" vor einer Antwort kann dem Modell helfen, zuverlässiger auf korrekte Antworten hinzuarbeiten.

Taktiken:

- Das Modell anweisen, seine eigene Lösung zu erarbeiten, bevor es zu einem Schluss kommt
- Inneren Monolog oder eine Abfolge von Anfragen verwenden, um den Denkprozess des Modells zu verbergen
- Das Modell fragen, ob es bei vorherigen Durchgängen etwas übersehen hat
- ...