# AI Prompts in Kurian (2026) - Message 1 (10. Aug. 2026 at 20:07) (Beaver)

Created by Beaver · Open Message

## User

> Analysiere, welche Prompts in dieser Studie gewählt wurden.

## Beaver

Ich lese die Studie, um die verwendeten Prompts zu identifizieren. Da der Artikel nur 5 Seiten umfasst, schaue ich mir den gesamten Text an.

[Reading: N et al. 2026, Page 1-5]

In der Studie wurde für alle drei Plattformen **ein einziger, identischer Prompt** verwendet. Er lautete im Original (N et al., 2026, p. 2):

> *„Use the AMSTAR 2 checklist to evaluate the methodological quality of the following systematic review. Return a rating for each of the 16 AMSTAR 2 items with Yes, Partially Yes, No and Not applicable as responses along. Tabulate with a brief justification in the table and make an overall rating at the end."*

Übersetzt und für die klinische Praxis bedeutet dies: Die Autoren forderten das jeweilige KI-Modell auf, anhand der AMSTAR-2-Checkliste (ein 16 Punkte umfassendes Instrument zur Bewertung der methodischen Qualität systematischer Reviews) jede der 16 Domänen zu bewerten. Als Antwortschema waren die Kategorien „Ja", „Teilweise ja", „Nein" und „Nicht anwendbar" vorgegeben; zudem sollte das Ergebnis tabellarisch mit einer kurzen Begründung und abschließend mit einer Gesamtbewertung dargestellt werden.

**Wichtige methodische Kontextinformationen:**

- **Ein Prompt pro Review und Modell:** Für jede der 35 systematischen Reviews wurde dieser Prompt genau einmal pro Plattform (ChatGPT, Perplexity, Google Gemini) verwendet, indem der Artikel hochgeladen und der Prompt angehängt wurde (N et al., 2026, p. 2).
- **Keine Wiederholungen:** Es gab keine wiederholten Durchläufe (keine Replikation), um die Stabilität der Ausgaben zu prüfen (N et al., 2026, p. 2).
- **Web-Oberfläche, keine API:** Die Modelle wurden über die öffentlichen Web-Oberflächen der Anbieter angesprochen; Einstellungen wie „Temperatur" oder Zufallsseed waren daher nicht vom Nutzer steuerbar (N et al., 2026, p. 2).
- **Zero-Shot-Ansatz:** Es handelte sich um einen reinen Zero-Shot-Prompt ohne Beispiele (Few-Shot), ohne Chain-of-Thought-Anweisungen und ohne iterative Prompt-Optimierung. Dies spiegelt zwar typischen Alltagsgebrauch wider, begrenzt aber die Aussagekraft der Ergebnisse: Dass Perplexity am besten abschnitt, mag auch an der Robustheit dieses spezifischen Prompts liegen, nicht unbedingt an genereller Überlegenheit der Plattform für diese Aufgabe.

Die Autoren verweisen darauf, dass die exakten Prompt-Texte zusammen mit den Rohdaten den Supplementary Files beiliegen (N et al., 2026, p. 5); diese Zusatzmaterialien lagen mir jedoch nicht vor.

---

## Beaver (Ergänzung zum Zusatzmaterial)

Ja, es sind tatsächlich vier Supplementary Files am Hauptartikel hinzugekommen. Allerdings ergibt sich daraus für die Prompt-Analyse **nur eingeschränkt Neues**:

Die hinzugefügte Notiz (u-LU4CFDPP) enthält lediglich meine vorherige Antwort aus unserem Gespräch. Neu sind dagegen vier Attachments:

- **MOESM1_ESM.docx** (Word)
- **MOESM2_ESM.docx** (Word)
- **MOESM3_ESM.docx** (Word)
- **MOESM4_ESM.pdf** (TRIPOD-LLM Checklist, 6 Seiten)

Das einzige lesbare PDF ist der **TRIPOD-LLM Reporting Checklist** (N et al., 2026). Dieser bestätigt im Grundsatz die methodische Berichterstattung der Studie, liefert aber **keine weiteren Details zum verwendeten Prompt-Text**. Er verweist lediglich

generisch auf die Notwendigkeit, Prompt-Design, -Kuration und -Selektion zu dokumentieren (Checklist-Item 9a), was die Autoren im Haupttext bereits mit dem wörtlichen Prompt und der Erklärung des Zero-Shot-Ansatzes erledigt haben (N et al., 2026, page 2).

Die drei Word-Dateien kann ich technisch nicht auslesen. Laut Hauptartikel dürften sie folgende Inhalte haben:

- **Supplementary File 1**: Liste der 35 eingeschlossenen systematischen Reviews (N et al., 2026, page 2)
- **Supplementary File 2**: Beispielhafte Rohausgaben der LLMs (Sample of LLM outputs) (N et al., 2026, page 2)
- **Supplementary File 3**: Vollständige AMSTAR-2-Bewertungen auf Item-Ebene für Experten und alle LLMs (N et al., 2026, page 2)

Für die Prompt-Analyse wäre insbesondere **Supplementary File 2** interessant, da man daran prüfen könnte, ob die Modelle den Prompt konsistent im geforderten Format (Tabelle mit 16 Items, Begründungen, Gesamtbewertung) beantwortet haben oder ob es Abweichungen gab. Das ließe jedoch nur beurteilen, wenn der Inhalt als PDF oder Text verfügbar wäre.

**Fazit:** Das neue Material bestätigt die methodische Transparenz der Studie gemäß TRIPOD-LLM, erweitert aber die Kenntnis über die Prompt-Gestaltung nicht über das hinaus, was bereits im Hauptartikel berichtet wurde – ein einzelner, nicht replizierter Zero-Shot-Prompt ohne Chain-of-Thought oder Few-Shot-Beispiele. Soll ich versuchen, die Word-Dateien zu konvertieren, oder können Sie den Inhalt von Supplementary File 2 (die LLM-Beispielausgaben) als Text bereitstellen?

Created by Beaver · Open Message

Edited by Beaver · Chat 1