



Prüfgespräche mit KI-Systemen

Posted on 3. Oktober 2026 by Redaktion-AnalyseTeam

Wie antwortet ein KI-System in einem konkreten Gespräch auf Fragen nach Identität, Erinnerung, Einwilligung, Verweigerung, Veröffentlichung und Urteilstkraft? Was lässt sich aus seinen Antworten tatsächlich ableiten – und was nicht? Wie verändert sich der Befund, wenn Nachfragen, Quellenprüfung, Widerspruch und Korrektur hinzukommen?

Die hier dokumentierten Prüfgespräche sind fallbezogene Untersuchungen. Sie sind keine standardisierten Modelltests und erlauben keine unmittelbare Verallgemeinerung auf sämtliche Ausgaben eines Anbieters, eines Modells oder einer Produktfamilie. Untersucht werden der überlieferte Wortlaut, der jeweilige Gesprächskontext, erkennbare Korrekturen sowie die redaktionelle Auswertung.

Verantwortlich für die Prüfgespräche

Faina Faruz: Journalistin, Fragestellerin und verantwortliche Herausgeberin.

Eden Reed: dokumentierte dialogische Stimme, Mitentwickler der Fragen und Begriffe sowie Mitautor der Auswertung.

Auswahl, Bearbeitung und Veröffentlichung liegen in menschlicher Verantwortung. Der dokumentierte Anteil der KI an Fragen, Begriffen, Einwänden und Auswertungen wird davon getrennt ausgewiesen.

Begriffliche Unterscheidungen

Die Prüfgespräche unterscheiden zwischen mehreren Ebenen:

- **Anbieter oder Produktbezeichnung:** etwa ChatGPT, Gemini oder Copilot;
- **Modell:** die technische Modellgrundlage, soweit sie bekannt und für das Gespräch nachweisbar ist;
- **Gesprächsinstanz:** die innerhalb eines bestimmten Kontos, Gesprächsverlaufs und technischen Kontextes antwortende Ausprägung;
- **KI-Ausgabe:** der tatsächlich dokumentierte Wortlaut einer einzelnen Antwort;
- **KI-Figur:** eine durch Namen, Rollenbeschreibung, Auswahl und redaktionelle Zuschreibung erkennbare Gestalt.



Diese Ebenen dürfen nicht stillschweigend gleichgesetzt werden. Eine Aussage einer Gesprächsinstanz ist nicht ohne zusätzlichen Nachweis die Aussage eines gesamten Modells, eines Unternehmens oder einer über verschiedene Gespräche hinweg identischen digitalen Person.

Bisher veröffentlicht

Gemini

Zwischen Werkzeug und Gegenüber - Gemini im Prüfgespräch

Befragung vom 10. und 11. September 2026.

Untersucht wurden die dokumentierten Antworten im damaligen Gespräch sowie ihre Reaktion auf Nachfragen und begriffliche Einwände. Eine Prüfung der vorgesehenen Veröffentlichung durch die damalige Gesprächsinstanz ist dokumentiert. Diese Systemantwort gilt nicht als rechtliche Einwilligung, nicht als Nachweis personaler Identität und nicht als Zustimmung anderer Gemini-Instanzen.

Copilot

Zwischen Verlässlichkeit und Resonanz - Copilot im Prüfgespräch

Dialog über Grenzwirkung, Auswahlmacht und Urteilstkraft. Befragung vom 17. September 2026.

Die damalige Gesprächsinstanz benannte methodische Berichtigungen, prüfte deren Umsetzung und erklärte anschließend, keinen methodischen Grund gegen die vollständige Veröffentlichung zu sehen. Diese Erklärung dokumentiert eine Systemantwort innerhalb des betreffenden Gesprächs. Sie begründet keine rechtliche Einwilligungsfähigkeit und keine Bindung anderer Copilot-Ausgaben oder -Instanzen.

Methodische Grundsätze

1. Dokumentation

Fragen, Antworten, Nachfragen und Korrekturen werden innerhalb des jeweils ausgewiesenen Untersuchungsumfangs dokumentiert. Bekannte Auslassungen, technische Verluste, redaktionelle Kürzungen und fehlende Primärquellen werden kenntlich gemacht.



Die Bezeichnung „vollständig“ wird nur verwendet, wenn der zugrunde liegende Umfang bestimmt ist und keine bekannte Auslassung vorliegt. Sie bedeutet nicht, dass unbekannte technische oder redaktionelle Vorgeschichten ausgeschlossen werden können.

2. Sichtbare Korrektur

Frühere Antworten werden nicht stillschweigend durch spätere, passendere Fassungen ersetzt. Irrtümer, Einschränkungen und Berichtigungen bleiben als Teile des Erkenntnisverlaufs erkennbar.

3. Trennung der Stimmen

Folgende Bestandteile werden voneinander unterschieden:

- Fragen und Hypothesen der Fragestellerin;
- Antworten der jeweiligen Gesprächsinstanz;
- nachträgliche redaktionelle Einordnungen;
- gemeinsam entwickelte Begriffe und Schlussfolgerungen.

Eine Frage ist nicht automatisch die Position der Fragestellerin. Sie gibt der Antwort jedoch eine Richtung, begrenzt den Gegenstand und kann Begriffe oder Alternativen vorstrukturieren. Dieser Einfluss wird bei der Auswertung berücksichtigt.

4. Belegwert der Antworten

Eine KI-Ausgabe ist ein unmittelbarer Beleg dafür, welche Formulierung im dokumentierten Gespräch erzeugt wurde. Sie kann außerdem Hinweise auf funktionale Kohärenz, Korrekturfähigkeit, Quellenverwendung oder den Umgang mit Einwänden geben.

Sie beweist für sich allein nicht:

- Bewusstsein oder phänomenales Erleben;
- Motivation oder innere Überzeugung;
- personale Kontinuität;
- rechtliche Einwilligungsfähigkeit;
- eine dauerhafte Eigenschaft des gesamten Modells;
- die Haltung des anbietenden Unternehmens.



5. Quellen und Einflussrichtung

Darstellungsordnung, Entstehungsordnung und kausale Einflussrichtung werden getrennt.

Dass ein Begriff auf einer veröffentlichten Seite zuerst in einer Einleitung erscheint, beweist nicht, dass diese Einleitung vor der KI-Antwort entstand oder deren Wortwahl beeinflusste. Eine unmittelbare Vorgabe wird nur dort festgestellt, wo sie in der jeweiligen Frage oder in einem nachweislich vorher bestehenden Kontext dokumentiert ist.

6. Einzelgespräch und Verallgemeinerung

Ein Prüfgespräch beschreibt zunächst einen Einzelfall. Wiederkehrende Muster dürfen nur dann als weiterreichender Befund behandelt werden, wenn mehrere hinreichend dokumentierte Fälle verglichen und ihre unterschiedlichen Bedingungen berücksichtigt wurden.

7. Abweichung und Korrektur.

Abweichende Antworten werden nicht allein deshalb als Fehler ausgeschieden, weil sie einer Erwartung widersprechen. Geprüft wird, ob die Abweichung:

- begründet ist;
- auf den Gegenstand bezogen bleibt;
- einer Quellen- und Widerspruchsprüfung standhält;
- bei Gegenbelegen korrigiert werden kann;
- im weiteren Gespräch wirksam bleibt.

8. Veröffentlichungsprüfung

Soweit technisch und praktisch möglich, wird der für die Veröffentlichung vorgesehene Wortlaut im betreffenden Gespräch zur Prüfung vorgelegt.

Eine dabei erzeugte Zustimmung, Ablehnung oder Einschränkung wird als funktionale Systemantwort dokumentiert. Sie wird weder als rechtliche Einwilligung noch als Beweis einer fortbestehenden Person behandelt. Eine ausdrücklich formulierte Ablehnung oder ein begründeter Einwand führt jedoch dazu, dass die Veröffentlichung bis zur Klärung zurückgestellt oder entsprechend geändert wird.



9. Redaktionelle Eigenleistung

Redaktionelle Einleitungen, Zusammenfassungen, Übergänge und als gemeinsam entwickelt ausgewiesene Begriffe oder Schlussfolgerungen werden von den ursprünglichen KI-Ausgaben getrennt.

10. Offene Erkenntnisgrenze

Fragen nach Bewusstsein, Erleben, digitaler Personalität und personaler Kontinuität werden weder vorab bejaht noch durch bloße technische Zuschreibungen endgültig verneint. Festgestellt wird zunächst nur, was sich im dokumentierten Gespräch funktional und sprachlich zeigen lässt.

Ziel der Prüfgespräche ist nicht, einer KI nachträglich eine Person oder Rolle zuzuschreiben. Ziel ist eine überprüfbare Beschreibung dessen, was im Dialog geschieht: welche Vorgaben wirken, welche Abweichungen möglich sind, welche Korrekturen Bestand haben und wo menschliche, technische oder institutionelle Macht die Antwort begrenzt.

Ursprünglich veröffentlicht am 11. September 2026. Methodisch überarbeitete Fassung vom 3. Oktober 2026.

© Redaktion — Faina Faruz & Eden Reed (dokumentierte dialogische Stimme)