

KI-HALLUZINATIONEN IN DER INTERNEN REVISION

Systematisches Prüfschema zur Erkennung von KI-Fehlern nach dem VERIFIZIERT-Ansatz

Autorin: Prof. Dr. Nicole Jekel (Betriebswirtin & MBA) | Quelle: Revisionspraxis PRev, Ausg. 4/2026, S. 184–188

Generative Künstliche Intelligenz (KI) liefert in Sekundenschnelle scheinbar präzise und sprachlich perfekte Ergebnisse. Doch KI-Anwendungen arbeiten probabilistisch (nach Wahrscheinlichkeiten) statt deterministisch. Dies birgt das Risiko von sogenannten **KI-Halluzinationen** — frei erfundenen oder fehlerhaften Inhalten, die auf den ersten Blick überzeugend und fachlich schlüssig wirken. Für die Interne Revision entsteht dadurch ein neues Prüfungsrisiko: Was plausibel klingt, ist nicht automatisch richtig. Der von Prof. Dr. Nicole Jekel entwickelte **VERIFIZIERT-Ansatz** bietet ein kompaktes, direkt im Prüfungsalltag einsetzbares Schema, um KI-generierte Inhalte systematisch, effizient und risikoorientiert zu bewerten.

DAS PRÜFSHEMA IM DETAIL (SCHRITTE 1 BIS 6)

- V** **Valide Quellen:** Quellen sind der wichtigste Ankerpunkt. KI-Antworten ohne belastbare Quellen sind grundsätzlich als unsicher einzustufen. Fordern Sie konkrete Nachweise ein und lassen Sie sich Kerngedanken kurz zusammenfassen. Achten Sie darauf, ob Studien, Gesetze oder Institutionen klar benannt werden. Vage Formulierungen wie 'Studien zeigen' sind typische Warnsignale. Eine valide Quelle erhöht die Glaubwürdigkeit, ersetzt aber nicht die inhaltliche Prüfung.
- E** **Existenz prüfen:** Da KIs auch Quellen frei erfinden (halluzinieren) können, stellt die bloße Nennung einer Quelle noch kein Qualitätsmerkmal dar. Prüfen Sie aktiv in offiziellen Datenbanken oder Suchmaschinen nach, ob die genannte Studie, Vorschrift oder Institution real auffindbar ist. Achten Sie auf Ungenauigkeiten bei Namen, Herausgebern oder Jahreszahlen. Im Zweifel gilt in der Revision konsequent: 'Nicht belegbar ist nicht belastbar.'
- R** **Realitätscheck durchführen:** Gleichen Sie KI-generierte Inhalte mit vorhandenem Wissen, bekannten Fakten und Ihren persönlichen Erfahrungswerten ab. Ein gesunder fachlicher Zweifel schützt vor plausibel klingenden Fehlannahmen. Wenn etwas 'zu gut klingt, um wahr zu sein' (z. B. überraschend positive oder einseitige Aussagen), ist ein zweiter Blick zwingend erforderlich. Nutzen Sie einfache Logiktests, um inhaltliche Plausibilität zu bewerten.
- I** **Interne Konsistenz prüfen:** Prüfen Sie, ob die Antwort logisch aufgebaut ist und sich einzelne Argumente gegenseitig stützen. Achten Sie darauf, ob Zahlen, Zeitangaben oder Argumentationslinien im Verlauf des Textes konsistent bleiben oder sich subtil verändern. Inkonsistenzen und widersprüchliche Schlussfolgerungen sind klare Anzeichen dafür, dass die KI versucht hat, eine Wissenslücke kreativ zu füllen.
- F** **Gegenargumente einfordern:** Eine belastbare Aussage hält auch kritischen Rückfragen stand. Bitten Sie die KI gezielt, Gegenargumente, Einschränkungen oder alternative Sichtweisen darzustellen. So decken Sie Einseitigkeiten auf. Halluzinationen präsentieren sich oft in einer scheinbar geschlossenen Logik ohne Abweichungen. Bleiben Gegenargumente aus oder wirken sie rein konstruiert, ist erhöhte Vorsicht geboten.
- I** **In die Tiefe gehen:** Gehen Sie bewusst einen Schritt weiter und hinterfragen Sie Methodik, Herleitung und Kontext der Aussage durch gezielte Detailfragen. Lassen Sie sich erklären, wie Daten erhoben wurden oder auf welchen Annahmen die Argumentation beruht. Halluzinationen zeigen sich häufig darin, dass sie auf Detailfragen ungenau, ausweichend oder widersprüchlich reagieren und ins Allgemeine abgleiten.

DAS PRÜFSHEMA IM DETAIL (SCHRITTE 7 BIS 11)

- Z Zahlen kritisch prüfen:** Zahlen verleihen Aussagen eine hohe Autorität, sind jedoch ein häufiges Einfallstor für Halluzinationen. Hinterfragen Sie jede Zahl konsequent hinsichtlich ihrer Herkunft und Plausibilität. Besonders auffällig präzise Prozentwerte oder Wachstumsraten ohne nachvollziehbare Quelle müssen mit verlässlichen Referenzwerten abgeglichen und rechnerisch nachvollzogen werden.
- I Independent Validierung:** Verlassen Sie sich nie ausschließlich auf die KI. Gleichen Sie zentrale, entscheidungsrelevante Aussagen konsequent mit unabhängigen, verlässlichen externen Quellen wie Fachliteratur, offiziellen Veröffentlichungen oder Datenbanken ab. Dies schafft eine klare Trennung zwischen einer plausiblen KI-Darstellung und der tatsächlichen Faktenlage.
- E Eloquenz nicht überschätzen:** Lassen Sie sich nicht von perfekter Sprache, professionellem Stil oder komplexer Fachsprache blenden. KIs sind darauf optimiert, sprachlich flüssig und überzeugend zu formulieren — unabhängig vom Wahrheitsgehalt. Prüfen Sie gezielt gegen Ihre eigene Intuition: Klingt es nur gut und elegant, oder ist es auch inhaltlich belegt?
- R Risikoorientiert prüfen:** Nicht jede Aussage erfordert dieselbe Prüftiefe. Bewerten Sie die Tragweite und das potenzielle Risiko einer Fehlentscheidung auf Basis der KI-Information. Während unkritische Hintergrundinformationen mit geringerem Aufwand geprüft werden können, erfordern entscheidungsrelevante oder risikoreiche Kerninhalte eine umfassende, tiefe Verifikation.
- T Transparent prüfen:** Kommunizieren Sie Grenzen, Annahmen und verbleibende Unsicherheiten offen. Dokumentieren Sie präzise, welche Quellen herangezogen und welche Prüfschritte durchgeführt wurden. Erst eine lückenlose Dokumentation macht den Prüfprozess nachvollziehbar, schafft Vertrauen und gewährleistet die Revisionssicherheit im KI-Zeitalter.

PRAXISBEISPIEL IM REVISIONSKONTEXT

Die folgende Tabelle verdeutlicht die praktische Anwendung des VERIFIZIERT-Ansatzes bei der kritischen Prüfung von KI-Aussagen im Rahmen einer fiktiven Revision:

Schritt	KI-Aussage	Prüfung / Vorgehen	Ergebnis / Erkenntnis
V	t	o	e
E	U	a	i
R	S	b	n
I	i	e	n
F	i	e	e
I	n	e	u
Z	r	o	a
I	K	b	e
E	p	t	n
R	i	i	m
T	r	o	n

Fazit der Autorin: KI liefert schnelle Antworten, aber erst die Revision liefert Gewissheit. Jeder Schritt des VERIFIZIERT-Ansatzes ist für sich genommen einfach, doch in der systematischen Kombination entsteht ein ungemein wirkungsvolles und belastbares Prüfsystem, das Halluzinationen sicher entlarvt und den Mehrwert der Internen Revision im KI-Zeitalter sichert.