



Sabine Anand und Birgit Spies

Grenzen generativer KI-Textproduktion im Lichte wissenschaftlicher Qualitätskriterien

Grenzen generativer KI-Textproduktion im Lichte wissenschaftlicher Qualitätskriterien

von Sabine Anand und Birgit Spies

Für wissenschaftliches Arbeiten haben sich zahlreiche Gütekriterien bewährt (vgl. z. B. Balzert, Schröder & Schäfer, 2011, S. 13–48; DFG, 2019; Döring, 2023, S. 79–118). Diese sind essenziell für die Entwicklung hochwertiger wissenschaftlicher Arbeiten von Fachexpert:innen und werden Studierenden bereits im ersten Semester vermittelt. Gleichzeitig werden Wissenschaftler:innen und Studierende auf die Möglichkeiten Künstlicher Intelligenz (KI) zur Unterstützung bei der Texterstellung hingewiesen. Insbesondere Generative Large Language Models (LLMs) wie ChatGPT werden zunehmend in der Wissenschaft und im Studium genutzt (Bašić, 2023). Warum auch nicht? Sie wirken auf den ersten Blick wie unerschöpfliche Wissensquellen und mit den „richtigen“ Prompts lassen sich schnell scheinbar plausible Texte generieren.

Dabei ist es allerdings wichtig zu verstehen, wie diese Tools funktionieren. Denn LLMs liefern nur auf Wahrscheinlichkeiten basierende Wortfolgen, die nicht reproduzierbar sind, kein eigenes Verständnis aufweisen und deren Herkunft nicht nachvollziehbar ist (vgl. z. B. James, 2024; Ranganathan, 2025). Genau diese Architektur führt auch dazu, dass zentrale Gütekriterien wissenschaftlichen Arbeitens systematisch verletzt werden – unabhängig davon, wie sogenanntes Prompt-Engineering, Retrieval-Augmentation oder „Verifizieren“ nachträglich eingesetzt werden.

Den aktuellen Diskurs unterstützend, betrachtet der vorliegende Beitrag die Grenzen wissenschaftlicher Texterstellung mittels generativer Künstlicher Intelligenz – insbesondere im Lichte wissenschaftlicher Qualitätskriterien, sinnvolle Möglichkeiten von LLM-Modellen in unterschiedlichen Kontexten ausdrücklich anerkennend.

Nachfolgend sei nun jedes Kriterium wissenschaftlichen Arbeitens einzeln beleuchtet und typische Fehlleistungen von Textproduktion mittels generativer Künstlicher Intelligenz aufgezeigt.

1. Objektivität

Objektivität bezieht sich auf die Unvoreingenommenheit und Sachlichkeit der Darstellung – eine Grundvoraussetzung wissenschaftlichen Arbeitens (Balzert, Schröder & Schäfer, 2011, S. 18 ff.). Dazu gehören sowohl der freie Zugriff auf Quellen als auch die kritische Reflexion eigener Vorurteile. LLMs hingegen spiegeln zum einen die Wertvorstellungen der Hersteller wider (Simon et al., 2024, S. 5). So können LLM-Akteure „definieren, welches Wissen [in] einer Kultur verfügbar ist und bleibt“ (Kerres et al., 2024, S. 1), was angesichts der herstellerseitigen Intransparenz der Arbeitsweise von LLMs nicht ausgeschlossen werden kann.

Zum anderen spiegeln LLMs die in den Trainingsdaten vorhandenen persönlichen und gesellschaftlichen Vorurteile wider (Simon et al., 2024, S. 8). Diese Vorurteile, sogenannte *Biases*, können nicht erkannt oder neutralisiert werden. LLMs übernehmen systematische Verzerrungen ihrer Trainingsdaten, wie etwa Geschlechter- oder Ethnizitätsstereotype, weil sie Wahrscheinlichkeiten aus vorhandenen Texten nachbilden (Deutscher Ethikrat, 2023, S. 373). Selbst moderne Debiasing-Ansätze reduzieren Fehlertreue nur partiell und können sogar neue Biases einführen (James, 2024). Da LLMs weder Intentionen noch ein Werturteil besitzen, fehlt echte Unvoreingenommenheit. Somit ist eine Objektivität der Ergebnisse nicht gegeben. Technologie ist nicht wertfrei.

2. Ehrlichkeit

Das Kriterium Ehrlichkeit umfasst die wahrheitsgemäße und aufrichtige Wiedergabe der gesammelten Daten, einschließlich einer kritischen Diskussion der Erkenntnisse, insbesondere dann, wenn diese unerwartet und/oder unerwünscht sind. Dies beinhaltet auch die Angabe aller genutzten Hilfsmittel und Quellen, einschließlich der Verwendung von KI-Tools, sowie aller beteiligten Personen. Damit sollen Plagiate, Täuschungen und/oder Datenmanipulationen vermieden werden (Balzert, Schröder & Schäfer, 2011, S. 15 ff.). Sogenannte *LLM-Halluzinationen* erzeugen jedoch frei erfundene, falsche, fehlerhafte oder irreführende Inhalte (Augenstein et al., 2024, S. 852). Je nach Anwendungsbereich lassen sich Fehlerquoten bis zu 83% feststellen (Garousi, 2025, S. 1; BBC, 2025, S. 2).

Die erzeugten Texte scheinen zwar sinnvoll, weichen aber z. T. deutlich von den Originalquellen ab (Siebert, 2024). Auch dieser Mechanismus ist architekturbedingt, weil das nächste sogenannte *Token* allein anhand statistischer Nähe gewählt wird, nicht anhand externer Wahrheitskriterien. Aufgrund dieser faktischen Fehler kann eine wahrheitsgemäße Datenwiedergabe nicht angenommen werden. Das Kriterium der Ehrlichkeit ist somit nicht erfüllt.

3. Überprüfbarkeit

Die konstruktive Kritik und die kritische Überprüfung der wissenschaftlichen Erkenntnisse ermöglichen Fehlerkorrekturen sowohl in der wissenschaftlichen Praxis als auch im Studium. Dies ist deshalb bedeutend, weil so schrittweise bessere Lösungen ermöglicht werden (Balzert, Schröder & Schäfer, 2011, S. 21 ff.). Daher ist es notwendig, das eigene Vorgehen nachvollziehbar zu beschreiben und Aussagen mithilfe wissenschaftlicher Quellen oder eigener, auf der Basis der Wissenschaftlichkeit erhobener, empirischer Daten zu belegen.

Neben fachlich falschen Inhalten (Garousi, 2025, S. 1; BBC, 2025, S. 2) erzeugen LLM-Modelle auch frei erfundene Quellenangaben (Walter & Wilder, 2023). Diese scheinen auf den ersten Blick realistisch, stellen sich aber bei einer Überprüfung als nicht existent heraus. So sind beispielsweise die Autoren, aber eben nicht der entsprechende Artikel zum Thema vorhanden. Auch weisen reale Zitate teils schwere Fehler auf (Siebert, 2024). Des Weiteren ist durch den eingangs erwähnten intransparenten Algorithmus von LLM-Modellen nicht nachvollziehbar, *welche* Quellen die Grundlage der Textgenerierung darstellen (Kerres et al., 2024, S. 2).

Zudem sind wissenschaftliche Zusammenfassungen und eine kritische Reflexion am Ende einer wissenschaftlichen Arbeit ebenfalls an das Kriterium der Überprüfbarkeit gebunden. Peters

und Chin-Yee (2025, S. 1) konnten zeigen, dass LLM-Modelle in 26% bis 73% der Fälle zu stark generalisieren, vereinfachen und fehlinterpretieren. Insgesamt entfällt die fundamentale Möglichkeit, Aussagen durch Primärquellen zu prüfen, auf einer überprüfbar verlässlichen Literatur- bzw. Datenbasis arbeiten sowie fachlich angemessene und nachvollziehbare Schlüsse ziehen zu können. Das Kriterium der Überprüfbarkeit ist nicht erfüllt.

4. Reliabilität

Reliabilität zeigt die Messgenauigkeit an und meint die Zuverlässigkeit der Messinstrumente und die Stabilität der Ergebnisse. Dies bedeutet, dass bei einer Messwiederholung die gleichen Ergebnisse erzielt werden (Balzert, Schröder & Schäfer, 2011, S. 26 ff.). Daher ist es wichtig, geeignete und passende Forschungsmethoden auszuwählen sowie die Instrumente vorher zu testen. Bei der Verwendung von LLMs hingegen können identische Prompts unterschiedliche oder sogar widersprüchliche Antworten liefern, weil es sich um wahrscheinlichkeitsabhängige Wortfolgen handelt (BSI, o. J.). Reliabilität ist somit nicht gegeben.

5. Validität

Mit Validität soll angezeigt werden, wie genau ein zu untersuchendes Merkmal geprüft wird und ob die Ergebnisse dementsprechend gültig sind. Dies bezieht sich z. B. auf ausreichend große, eventuell repräsentative Stichproben, die Bearbeitung der richtigen Inhaltsbereiche sowie die Passgenauigkeit der Fragen (Balzert, Schröder & Schäfer, 2011, S. 27 ff.). Auch hier ist eine kritische Überprüfung daraufhin notwendig, ob tatsächlich gemessen wird, was gemessen werden sollte. LLMs beantworten Fragen zwar scheinbar plausibel, aber ohne echten Konstruktbezug. Bei Formulierungsänderung bricht die Leistung ein, insbesondere bei mehrstufiger Deduktion und untersuchten Merkmalen, die praktisches Wissen erfordern (Boye & Moëll, 2025). Validität kann damit nicht durchgängig angenommen werden.

6. Verständlichkeit

Dies bezieht sich auf die Zweckmäßigkeit, die Folgerichtigkeit sowie die Vollständigkeit einer wissenschaftlichen Arbeit. Dazu gehören z. B. die inhaltliche Struktur, die Bestandteile der Arbeit, die sprachliche Aufbereitung und das Erscheinungsbild (Balzert, Schröder & Schäfer, 2011, S. 29 ff.). In wissenschaftlichen Arbeiten ist es für Fachexpert:innen und Studierende umso mehr von Bedeutung, eine klare und präzise Sprache zu verwenden (vgl. „Verständlichkeitsmacher“ von Schulz von Thun, 2013) und alle relevanten Fachbegriffe zu definieren. Obwohl die generierten Texte von LLMs sprachlich flüssig erscheinen, mangelt es oft an Kohärenz. Dabei führen insbesondere längere Konversationen zu Kontextverlust. Es kann vorkommen, dass das LLM-Modell frühere Aussagen „vergisst“ und sich selbst widerspricht. Zudem werden Fachtermini teilweise inflationär eingesetzt, ohne Begriffe klar zu definieren und begriffliche Beziehungen korrekt abzubilden. Die sprachliche Flüssigkeit täuscht Folgerichtigkeit und Verständlichkeit vor (Ranganathan, 2025). Die Kriterien der wissenschaftlichen Verständlichkeit und Folgerichtigkeit werden nicht erfüllt.

7. Logische Argumentation

Logische Argumentation setzt ein klares Verständnis der zugrundeliegenden theoretischen Konstrukte und Fachbegriffe voraus. Die Verwendung passender Argumentationstypen ist entscheidend, um Schlussfolgerungen zu ziehen, diese entsprechend zu untermauern und letztendlich Fehlschlüsse zu vermeiden. Folgerichtige und widerspruchsfreie Begründungen der Aussagen lassen sich beispielsweise in einem gemeinsamen Gespräch überprüfen. Doch LLM-Modelle „denken“ anders. Formale Logik-Tests, wie die Wahlaufgabe von *Wason* (Beyer & Gerlach, 2018, S. 104), zeigen, dass LLMs Denkfehler (re-)produzieren. Auch werden zu verschiedenen Zeiten auf dieselbe Frage unterschiedliche Antworten gegeben (Macmillan-Scott & Musolesi, 2024). Fehler beim konditionalen Schlussfolgern entstehen, weil die *Token-Prädiktion*, also die Wahrscheinlichkeitsvorhersage für das nächste Wort, keine symbolische Ableitungsfähigkeit besitzt. Dabei wird, entsprechend der Bestätigungstendenz, eher nach bestätigender und nicht nach falsifizierender Information gesucht. „*Was im ersten Moment logisch klingt, muss nicht logisch sein.*“ (Balzert, Schröder & Schäfer, 2011, S. 34). Das Kriterium der logischen Argumentation kann nicht als durchgängig erfüllt angenommen werden.

8. Relevanz

Der Informationswert der Ergebnisse ist ein weiteres bedeutendes Kriterium wissenschaftlichen Arbeitens. In der Wissenschaft ist relevant, „*was im Fachgebiet neues Wissen schafft [...] [und] zum wissenschaftlichen Fortschritt beiträgt*“ (Balzert, Schröder, Schäfer, 2011, S. 32). Relevant und bedeutsam sind beispielweise neue Produkte und Verfahren, welche helfen, Praxisprobleme zu lösen. Auch wenn LLMs in Sekunden große Texte generieren können, erschaffen Sie dennoch nicht NEUES Wissen. Im Gegenteil: irrelevante, irreführende oder thematisch abweichende Inhalte, welche massenhaft durch LLMs selbst generiert werden, überfluten Suchmaschinen, Preprint-Server und Peer-Review-Kanäle. „*[D]ie Grenze zwischen Wahrheit und Lüge [verschwindet] hinter plausibel klingenden Texten*“ (Simon et al. 2024, S. 12; n. Hicks et al. 2024, S. 1). Denn, wie bereits oben ausgeführt, reproduziert generative Künstliche Intelligenz gerade durch ihren auf Wahrscheinlichkeiten beruhenden Algorithmus „*das bestehende Wissen, die bestehende Weltsicht, [...] und verhärtet sie [...]. [...] und verschließt sich damit vor jedweder Entwicklung*“ (Limburg et al. 2023, S. 15). Mittels generativer KI erstellte Texte erhöhen somit letztlich die Menge nicht-relevanter Treffer und erschweren die gezielte Literaturrecherche (Zhou et al., 2023). Das Kriterium der Relevanz wird nicht erfüllt.

9. Originalität

Das Kriterium Originalität bezieht sich auf die Eigenständigkeit der Leistungen. Wissenschaftliche Arbeiten sollen originale und neue Lösungen hervorbringen. Heuristische Suchstrategien und zahlreiche Kreativitätsmethoden, wie z. B. Perspektivwechsel und Brainstorming sowie verschiedene systematisch-analytische Methoden können hier angewendet werden (Balzert, Schröder & Schäfer, 2011, S. 39 ff.). Dies im Rahmen des Studiums zunächst zu erlernen und anschließend zu vertiefen, ist Bestandteil des Studiums selbst. Werden im Arbeits- und Lernprozess die Erarbeitungs-, Vertiefungs- und

Anwendungsphase durch automatisierte Möglichkeiten übersprungen, so können ebendiese Fähigkeiten und Fertigkeiten nicht ausgebildet und verbessert werden und letztlich zu *deskilling* (Reinmann 2023, S.4 ff.) und *skillskipping* (Nuxoll, 2024) führen. Sofern das Kriterium der Originalität als Gütekriterium für die Anfertigung wissenschaftlicher Arbeiten weiter Bestand haben soll, ist die Nutzung von LLM-Modellen daher kritisch zu sehen. Sie widerspricht zum einen dem Prinzip der Eigenständigkeit und führt zum anderen zu einem Verlust von Originalität wissenschaftlicher Texte (Limburg et al., S. 11). LLMs erzeugen Paraphrasen, die gängige Plagiat-Checker umgehen können (Lee et al., 2025). Dadurch können verdeckte Urheberrechtsverletzungen entstehen. Die Nutzung von KI-Tools muss daher angezeigt werden. Die Frage nach Originalität bleibt dennoch bestehen (Kleebayoon & Wiwanitkit, 2023; Hutson, 2024).

10. Nachvollziehbarkeit

Hier ist es bedeutend, inwieweit bereits erwähnte Qualitätskriterien erfüllt werden, wie z. B. objektive und sachliche Beschreibungen, verständliche Darstellung und Strukturierung von Inhalten mit hohem Informationsgehalt, richtiges Sampling und die passenden Gültigkeitsbereiche, fehlerfrei arbeitende Messinstrumente, welche zu überprüfbaren Aussagen führen sowie widerspruchsfreie Argumentation und logisch einwandfreie Schlussfolgerungen. Die Entscheidungspfade von LLM-Modellen sind nicht einsehbar, nicht transparent und verhindern so die Nachvollziehbarkeit der Vorgehensweise (Simon et al. 2024, S. 12, Deutscher Ethikrat, 2023, S. 376). Das Kriterium der Nachvollziehbarkeit ist nicht gegeben.

11. Fairness

„Zur Fairness gehören Ehrlichkeit, Unparteilichkeit, Kollegialität, gegenseitiger Respekt und die ehrliche Anerkennung der Leistung anderer Personen“ (Balzert, Schröder & Schäfer, 2011, S. 45). Wie bereits erwähnt, werden diskriminierende Muster durch LLMs reproduziert. Zudem sind Inhalte und Quellenangaben teilweise frei erfunden, falsch, fehlerhaft oder irreführend (Simon et al., 2024, S. 13 ff.; Hicks et al., 2024). Die Leistung anderer Personen kann durch diese Urheberrechtsverletzungen nicht anerkannt werden. Fairnessziele können daher mit der Verwendung von LLMs gar nicht erreicht werden. *„[...] faire Wissenschaft [ist] kein finales Produkt, sondern gelebte Praxis [...], die durch stetiges Hinterfragen der bestehenden Strukturen immer wieder neu verhandelt werden muss.“* (Frisch, Hagenström & Reeg, 2022, S. 242).

12. Verantwortung

Dieses Kriterium schließt die Verantwortung gegenüber sich selbst, gegenüber anderen sowie die Verantwortung gegenüber der Wissenschaft und der Gesellschaft im Allgemeinen ein. Zudem bezieht es sich auf bewusstes Ausrichten des eigenen Handelns nach ethischen Maßstäben. LLM-Modelle besitzen keine Rechtspersönlichkeit. Ohne klare Zurechenbarkeit bleibt wissenschaftliche Verantwortung unerfüllbar. Wer beispielsweise *„kritische Reflexion vollständig an Maschinen abgibt, überträgt den Maschinen auch die Verantwortung für die*

Schlüsse, die aus wissenschaftlichen Erkenntnissen gezogen werden sollten.“ (Limburg et al. 2023, S. 20). Die Einhaltung wissenschaftlicher, fachlicher und ethischer Maßstäbe im Rahmen einer verantwortungsvollen KI-Nutzung sollte jedoch beim Nutzer bleiben.

Fazit

Dieser Beitrag hat Grenzen generativer KI-Textproduktion im Lichte wissenschaftlicher Qualitätskriterien zusammengefasst. Dabei wurde gezeigt, dass Tools der generativen Künstlichen Intelligenz zur Erstellung wissenschaftlicher Texte aufgrund ihrer Funktionsweise essenzielle wissenschaftliche Qualitätsstandards verfehlen, welche aber die Grundlage wissenschaftlichen Erkenntnisgewinns darstellen. Durch reine KI-Generierung können daher keine hochwertigen wissenschaftlichen Arbeiten entstehen.

Zum einen ist es für Wissenschaftler:innen und die wissenschaftliche Arbeit selbst unerlässlich, Quellen vertrauen und als Basis für die eigene Arbeit heranziehen zu können. Sowohl Wissenserwerb als auch wissenschaftlicher Erkenntnisgewinn bedürfen eines verlässlichen Fundaments.

Zum anderen ist es für Studierende, die sich in der Phase des Wissenserwerbs befinden, unumgänglich, sich notwendige Techniken, wie die eigene Texterstellung und die Auseinandersetzung mit wissenschaftlichen Texten, mittels derer der eigene geistige Horizont erweitert werden kann, anzueignen. *„Denn Schreiben ist nicht nur Medium zur Textproduktion, sondern auch Instrument des Lernens, zur Erkenntnisgewinnung sowie zur präzisen und effektiven Kommunikation von Erkenntnissen“* (Limburg et al. 2023, S. 3 f.). Insbesondere sind hier kognitive Fähigkeiten entsprechend der Taxonomie-Stufen nach Bloom et al. (1973), auszubilden. Diese sind u. a. in den Qualifikationsprofilen der Studiengänge beschrieben und umfassen wissen, verstehen, anwenden, analysieren, beurteilen und erschaffen (Bloom et al., 1973). Limburg et al. (2023, S. 14 f.) weisen darauf hin, dass besonders kritisches Denken leicht ins Hintertreffen geraten und verkümmern kann - die Fähigkeit zu analysieren, zu bewerten, kritisch zu hinterfragen und, darauf aufbauend, neue Gedankengänge und Lösungsmöglichkeiten zu entwickeln. Auch Gerlich (2025, S. 1) führt aus, dass das Vermögen kritisch zu denken mit der Häufigkeit der Nutzung von KI-Tools abnimmt. Ebenso vermindern Nutzer in der Folge die eigene Anstrengung in vertiefte und reflektierte Denkprozesse (*cognitive offloading*) (Gerlich, 2025, S. 24).

Reinmann (2023) schlägt als gedankliche Ausgangspunkte für konkrete Veränderungen in der wissenschaftlichen Arbeit und dem Hochschulkontext Folgendes vor: *„[1] eine Renaissance der menschlichen Vernunft als Gegengewicht zur künstlichen Intelligenz, [2] eine Aufwertung sozialer Beziehungen als Gegengewicht zur Mensch-Maschine Relation und [3] einen radikalen Wandel der Prüfungskultur als Gegengewicht zu juristischen Absicherungen“* (Reinmann 2023, S. 15). Hier kann angesetzt und im Interesse eines wissenschaftlichen Erkenntnisgewinns weitergearbeitet werden.

Die Integration generativer KI in die Hochschullandschaft erfolgt derzeit weitgehend ungeplant und reaktiv. Anstatt kritisch-reflektierte Gestaltung zu betreiben, reagieren viele Institutionen lediglich auf bereits eingetretene Veränderungen – ein Ansatz, der erhebliche Risiken für die wissenschaftliche Integrität und Qualität birgt, wie in diesem Beitrag gezeigt wurde. Der Einsatz generativer KI bei der Erstellung wissenschaftlicher Texte erfordert eine kritische Begleitung mit systematischen und multiperspektivischen Ansätzen. Hier sollten sowohl die technologischen Möglichkeiten, die Lehr- und Lernpraxis an Hochschulen als auch die ethischen und wissenschaftlichen Herausforderungen berücksichtigt werden. Dazu gehören z. B. die kritische Neugestaltung von Prüfungsformaten, die Überarbeitung von Lehr- und Lernzielen und Curricula zur Kompetenzentwicklung für Lehrende und Studierende sowie eine kontinuierliche Evaluation und Anpassung. Weitere Forschung dazu ist dringend empfohlen.

Literatur

- Augenstein, I. et al. (2024). Factuality challenges in the era of large language models and opportunities for fact-checking. In *Nature Machine Intelligence*, 6, S. 852-863. Verfügbar unter: <https://www.nature.com/articles/s42256-024-00881-z> [29.08.2025].
- Balzert, H., Schröder, M. & Schäfer, C. (2011). *Wissenschaftliches Arbeiten. Ethik, Inhalt & Form wiss. Arbeitens, Handwerkszeug, Quellen, Projektmanagement, Präsentation* (2. Aufl.). Herdecke Witten: W3L.
- Bašić, Z., Banovac, A., Kružić, I. & Jerković, I. (2023). ChatGPT-3.5 as writing assistance in students' essays. In *Humanities and Social Sciences Communication*, 10, Art. Nr. 750. Verfügbar unter: <https://doi.org/10.1057/s41599-023-02269-7> [29.08.2025].
- BBC (2025). *Representation of BBC News content in AI Assistants*. Verfügbar unter: <https://www.bbc.co.uk/aboutthebbc/documents/bbc-research-into-ai-assistants.pdf> [29.08.2025].
- Beyer, R. & Gerlach, R. (2018). *Sprache und Denken* (2. Auflage). Wiesbaden: Springer.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1973). *Taxonomie von Lernzielen im kognitiven Bereich* (3. Aufl.). Weinheim, Basel: Beltz.
- Boye, J. & Moëll, B. (2025). *Large Language Models and Mathematical Reasoning Failures*. Verfügbar unter: <https://doi.org/10.48550/arXiv.2502.11574> [29.08.2025].
- BSI - Bundesamt für Sicherheit in der Informationstechnik (Hrsg.) (o. J.). *Texte & Co. mit Künstlicher Intelligenz erstellen - Generative KI und ihre Risiken*. Bonn: Bundesamt für Sicherheit in der Informationstechnik. Verfügbar unter: https://www.bsi.bund.de/DE/Themen/Verbraucherinnen-und-Verbraucher/Informationen-und-Empfehlungen/Technologien_sicher_gestalten/Kuenstliche-Intelligenz/KI-Texterstellung/ki-texterstellung.html [29.08.2025].
- Deutscher Ethikrat (Hrsg.) (2023). *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*. Stellungnahme. Berlin: Deutscher Ethikrat. Verfügbar unter: <https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf> [29.08.2025].
- DFG - Deutsche Forschungsgemeinschaft (Hrsg.) (2019). *Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex*. Bonn: Deutsche Forschungsgemeinschaft e. V. Verfügbar unter: https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf [29.08.2025].
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. Aufl.). Berlin, Heidelberg: Springer.
- Frisch, K.; Hagenström, F. & Reeg, N. (2022). *Wissenschaftliche Fairness. Wissenschaft zwischen Integrität und Fehlverhalten*. Bielefeld: transcript. Verfügbar unter: <https://d-nb.info/1270403435/34> [29.08.2025].
- Garousi, V. (2025). *Why you shouldn't fully trust ChatGPT: A synthesis of this AI tool's error rates across disciplines and the software engineering lifecycle*. Verfügbar unter: <https://arxiv.org/pdf/2504.18858> <https://doi.org/10.48550/arXiv.2504.18858> [29.08.2025].

- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. In *Societies* 2025, Vol. 15(1), 6. Verfügbar unter: <https://doi.org/10.3390/soc15010006> [29.08.2025].
- Hicks, M., Humphries, J. & Slater, J. (2024). Chat GPT is bullshit. In *Ethics and Information Technology*, Vol. 26, Art. Nr. 38. Verfügbar unter: <https://doi.org/10.1007/s10676-024-09775-5> [29.08.2025].
- Hutson, J. (2024). Rethinking Plagiarism in the Era of Generative AI. In *Journal of Intelligent Communication*, Vol. 4 (1). Verfügbar unter: <https://digitalcommons.lindenwood.edu/cgi/viewcontent.cgi?article=1626&context=faculty-research-papers> [29.08.2025].
- James, M. (2024). Understanding LLM Errors: What They Are and How to Address Them. In *Dev Community Blog*. Verfügbar unter: <https://dev.to/thenexttech/understanding-llm-errors-what-they-are-and-how-to-address-them-42f9> [29.08.2025].
- Kerres, M., Klar, M., Mulders, M. (2024). Informationskompetenz neu denken: Von Google zu ChatGPT. In *Erwachsenenbildung*, Vol. 70(2), S. 52-57. Verfügbar unter: <https://www.researchgate.net/publication/378776677> Informationskompetenz neu denken Von Google zu ChatGPT [29.08.2025].
- Kleebayoon A. & Wiwanitkit, V. (2023). Artificial intelligence, chatbots, plagiarism and basic honesty: comment. In *Cellular and molecular bioengineering*, 2023-04, Vol.16 (2), S.173-174. Verfügbar unter: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10121937/> [29.08.2025].
- Lee, J., Agrawal, T., Uchendu, A., Le, T., Chen, J. & Lee, D. (2024). *PlagBench: Exploring the Duality of Large Language Models in Plagiarism Generation and Detection*. Verfügbar unter: <https://doi.org/10.48550/arXiv.2406.16288> [29.08.2025].
- Limburg, A., Bohle-Jurok, U., Buck, I., Grieshammer, E., Gröpler, J., Mundorf, M., Schindler, K., Wilder, N. (2023). Zehn Thesen zur Zukunft des Schreibens in der Wissenschaft. In *Hochschulforum Digitalisierung*, Diskussionspapier 23. Verfügbar unter: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_DP_23_Zukunft_Schreiben_Wissenschaft.pdf [29.08.2025].
- Macmillan-Scott, O. & Musolesi, M. (2024). (Ir)rationality and cognitive biases in large language models. In *Royal Society Open Science*, Vol. 11 (6): 240255. Verfügbar unter: <https://doi.org/10.1098/rsos.240255> [29.08.2025].
- Nuxoll, F. (2024). Skill Skipping. Was Lehrpersonen tun können, damit KI nicht Lernprozesse verhindert. In *Lehren & Lernen* Vol. 50 (6/2024), S. 11-12. Verfügbar unter: https://www.pedocs.de/volltexte/2025/32580/pdf/LehrenundLernen_2024_6_Nuxoll_Skill_Skippping_Was_Lehrpersonen.pdf [29.08.2025].
- Peters, U. & Chin-Yee, B. (2025). Generalization bias in large language model summarization of scientific research. In *Royal Society Open Science*, Vol. 12 (4): 241776. Verfügbar unter: <https://doi.org/10.1098/rsos.241776> [29.08.2025].
- Ranganathan, B. (2025). Limitations of LLM Reasoning. In *DZone Blog*, verfügbar unter: <https://dzone.com/articles/llm-reasoning-limitations> [29.08.2025].
- Reinman, G. (2023). Deskillung durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik. In *Hochschulforum Digitalisierung*, Diskussionspapier Nr. 25. Verfügbar unter:

https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_DP_25_Deskilling.pdf [29.08.2025].

Schulz von Thun, F. (2013). *Miteinander reden 1: Störungen und Klärungen: allgemeine Psychologie der Kommunikation* (50. Aufl.). Reinbek bei Hamburg: Rowohlt.

Siebert, J. (2024). Halluzinationen von generativer KI und großen Sprachmodellen (LLMs). In *Blog des Fraunhofer-Institut für Experimentelles Software Engineering*. Verfügbar unter: <https://www.iese.fraunhofer.de/blog/halluzinationen-generative-ki-llm/> [29.08.2025].

Simon, J., Spiecker gen. Döhmman, I. & von Luxburg, U. (2024). *Generative KI – jenseits von Euphorie und einfachen Lösungen. Diskussion Nr. 34*. Halle (Saale): Nationale Akademie der Wissenschaften Leopoldina. Verfügbar unter: https://doi.org/10.26164/leopoldina_03_01226 [29.08.2025].

Walters, W. H. & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. In *Scientific Reports* 13, Art. Nr. 14045. Verfügbar unter: <https://doi.org/10.1038/s41598-023-41032-5> [29.08.2025].

Zhou, J., Zhang, Y., Luo, Q., Parker, A. G. & De Choudhury, M. (2023). *Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions*. CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Art. Nr. 436, S. 1 – 20. Verfügbar unter: <https://doi.org/10.1145/3544548.3581318> [29.08.2025].

Über die Autorinnen



Prof. Dr. Sabine Anand, Professorin für Bildungswissenschaften an der Hochschule Fresenius in Köln

Sabine Anand studierte Erziehungswissenschaft, Germanistik sowie Kommunikations- und Medienwissenschaft (M. A.) und promovierte im Fach Erziehungswissenschaft an der Universität Leipzig in Kooperation mit der Universität von Kapstadt, Südafrika (Wirtschaftswissenschaft) zum Dr. phil. Sie ist im Fernstudium Studiendekanin für den Studiengang Kindheitspädagogik, Leitung und Management (B. A.).

Kontakt: sabine.anand@hs-fresenius.de



Prof. Dr. Dipl.-Ing. Birgit Spies, Professorin für Bildung und Digitalisierung an der Hochschule Fresenius in Hamburg

Birgit Spies studierte Informationstechnik in Dresden (Dipl.-Ing.), Medien und Bildung in Rostock (M.A.) und promovierte an der LMU in München in den Fächern Pädagogik (H) und Psychologie (N) zum Dr. phil. Ihre Arbeitsschwerpunkte sind Medienbildung, Mediendidaktik und Medienpsychologie. Sie ist im Fernstudium Studiendekanin der Studiengänge Erwachsenenbildung (M.A.), Bildung, Medien und Digitalisierung (M.A.) sowie Bildung, Management und Führung (M.A.).

Kontakt: birgit.spies@hs-fresenius.de

Impressum

CC BY NC ND – Sabine Anand & Birgit Spies, 2025



Dieser Beitrag ist unter einer Creative-Commons-Lizenz verfügbar. Der Download und die Weitergabe dieser Publikation unter Nennung des Namens und der Quelle ist ausdrücklich erwünscht. Jegliche Bearbeitung oder kommerzielle Nutzung des Inhalts ist ausgeschlossen.

<https://creativecommons.org/licenses/>

Herausgabe: Open Access Journal „Bildung – Mensch – Medien“

<https://bildung.hypotheses.org>

ISSN: 2942-4224

Titelbild: #2149314038 | freepik.com

Coveridee: Anja Lindenberg

Empfohlene Zitierweise

Anand, S., Spies, B. (2025). Grenzen generativer KI-Textproduktion im Lichte wissenschaftlicher Qualitätskriterien. In *Bildung-Mensch-Medien 2025*. Verfügbar unter: <https://bildung.hypotheses.org/7188> [Datum des Zugriffs].