

Generative KI wie ChatGPT als Lehr- und Lernhilfe in der Hochschullehre

Roman Bartnik

Invalid Date

Table of contents

1 Generative KI wie ChatGPT als Lehr- und Lernhilfe in der Hochschullehre	3
2 Kapitel 1 – KI als Hilfe zum Lehren und Lernen	4
3 Kapitel 2 – Zentrale Komponenten großer Sprachmodelle	9
4 Kapitel 3 – Ziele und didaktische Mechanismen - was wollen wir mit KI erreichen? 10	
5 Kapitel 4 – Vier Szenarien: Hiwi, Copilot, Tutor, Simulator	11
6 Kapitel 5 – Empfehlungen zur Umsetzung	12
7 Appendix 1 – Risiken und Nebenwirkungen: Halluzinierende Bots, Betrug und Desinformation	13
8 Appendix 2 – Beispiele für KI-Richtlinien	14
9 Appendix 3 – Ausgewählte Prompts zur Lehr- und Lernunterstützung	15
10 Appendix 4 – Detaillierte Beispiele für Aufgabenstellungen mit KI-Unterstützung	16
	17

1 Generative KI wie ChatGPT als Lehr- und Lernhilfe in der Hochschullehre

Skript zum Workshop, © Roman Bartnik, TH Köln. Version 8, 2025/08/26 , roman.bartnik@th-koeln.de (Updates zur letzten größeren Vorversion (V5):. (1), (4): Umfassende Einbindung neuer Studien und Anwendungsbeispiele; Anmerkungen zu **GPT-5**. (2): Ergänzung von neuen Visualisierungen der LLM Denkprozesse in (2.2.) und Diskussion von neuen kostenfreien **Angeboten für Lehrende** wie Academic Cloud u.a. (2.3). Neue empirische Studien zur Wirksamkeit von Prompt-Strategien (2.5.2). Ergänzung eines Abschnittes zum **Energieverbrauch der KI-Modelle** (2.6).(4.4.) Ergänzung neuer Anwendungsbeispiele für KI als Simulator. (5.4.): Ergänzung eines Abschnitts zum **Prüfen unter KI Bedingungen**. (5.5.) Übersicht zu **Richtlinien der KI Nutzung**.

2 Kapitel 1 – KI als Hilfe zum Lehren und Lernen

Wie kann uns **generative künstliche Intelligenz (KI)** in der Lehre helfen? Hoffnung besteht hier für zwei typische Probleme: Erstens haben Studierende **individuelle Bedürfnisse**, aber wir haben nur **begrenzte Zeit**, auf diese einzugehen. Wie können wir Einzelne möglichst intensiv fördern, ohne vor Arbeit unterzugehen? Zweitens ist der **Aufwand gerade für effektive Lehrmethoden** oft sehr hoch – so etwa für häufige niedrigschwellige Tests oder **individuelles Feedback zu Studienarbeiten** (s. etwa Hattie (2023), Kap. 13; Brown, Roediger, and McDaniel (2014)). Wer lehrt, fühlt sich aus Zeit- und Stoffdruck oft gezwungen, Abstriche von idealen Lehr-Setups zu machen (Henderson and Dancy (2007); Schmidt and Tippelt (2005), S. 104–105). Gerade Lehrmethoden, die didaktisch sinnvoll, aber mit hohem Aufwand verbunden sind, drohen dabei auf der Strecke zu bleiben (Brown, Roediger, and McDaniel (2014)).

Für die Lehre erschließen sich durch die großen KI-Sprachmodelle (LLM = Large Language Models) neue Möglichkeiten. Sie sind – wie es eine Analyse des MIT-Professors Andrew McAfee auf den Punkt bringt – „**generally faster**“ (McAfee (2024)). Lehrende können mit KI-Unterstützung etwa deutlich schneller ein Set von Übungsaufgaben erstellen, mehrere Anwendungsbeispiele pro Konzept hinzufügen, Quizfragen zur schnellen Lernüberprüfung generieren oder mit den Studierenden Rollenspiele durchführen (Meincke, Mollick, and Terwiesch (2024); Mollick and Mollick (2023)).

Der Berg ist noch da, aber mit dem E-Bike kommt man weiter.

Immer mehr Aspekte typischer Forschungstätigkeiten – ein zentraler Ausbildungsinhalt der Hochschulen – können **zunehmend von der KI übernommen** werden, und zwar auf hohem Niveau. Vorbei sind die Zeiten, in denen wir die banalen Schreibprodukte der KI nur belächeln konnten. Ein Überblicksartikel des Forschers Anton Korinek im renommierten *Journal of Economic Literature* vom Dezember 2024 fasst das deutlich höhere Niveau zusammen:

„Die derzeitige Generation von LLMs ist in hohem Maße in der Lage, die wichtigsten Erkenntnisse von Forschungsarbeiten zu verarbeiten.“
(Korinek (2024), S. 3; Übersetzung RB mit DeepL)

Die professionelle Nutzung ist hier noch weiter: So demonstrierte etwa Google (2025) ein mehrstufiges Modell für die Pharma-Forschung („**AI co-scientist**“), das den Forschenden zeitintensive Zwischenschritte abnimmt (Gottweis et al. (2025)). Auch im Peer-Review werden

zunehmend Sprachmodelle eingesetzt – mit allen Vor- und Nachteilen, die das mit sich bringt (Naddaf (2025a)). Wie wir in den späteren Kapiteln sehen, experimentieren Hochschulen weltweit intensiv mit den neuen Möglichkeiten für Lehre und Forschung.

Drei zentrale Weiterentwicklungen zwischen 2024 und 2025 sind laut Korinek (2024) (S. 2–3) für den deutlichen Sprung in forschungsrelevanten Fähigkeiten der Sprachmodelle verantwortlich:

Erstens **neue Interaktionsmöglichkeiten** - Während die typische Nutzung früher auf Texteingabe im Eingabefenster beschränkt war, bieten die großen Sprachmodelle mittlerweile die Möglichkeit, in einem Workspace gemeinsam an Text oder Code zu arbeiten (z. B. ChatGPT Canvas, Claude Artifacts).

Zweitens eine deutliche Verbesserung der **Problemlösefähigkeit (Reasoning)** der Modelle. Den stärksten Modellen (GPT-5, Gemini 2.5, Claude Opus 4.1) kann man mittlerweile dabei zusehen, wie sie mehrstufiges Problemlösen und logisches Schlussfolgern etwa bei Rechercheaufgaben durchführen. Die Bedeutung präziser Prompt-„Zaubersprüche“ nimmt ab, da die neueren Reasoning-Modelle ohnehin selbst Schritt für Schritt vorgehen und nachfragen (Meincke et al. (2025a)). Insgesamt steigt seit 2023 die Qualität der Aufgaben, die Sprachmodelle erledigen können, stark an. Empirische Untersuchungen zeigen, dass die Modelle immer längere Aufgaben auf hohem Niveau bearbeiten können (Kwa et al. (2025)).

Was ändert sich durch GPT-5? Aus User-Sicht ist GPT-5 im Vergleich zu Vormodellen selbstständiger geworden – User müssen nicht mehr selbst zwischen vielen unterschiedlichen Modellen auswählen. Je nachdem, wie einfach die Frage ist, wird Schnelligkeit bevorzugt (durch Nutzung eines kleineren Modells wie GPT-5 nano) oder es wird ein schwereres Werkzeug angelegt (mehrstufiges Suchen und Reflektieren mit einem größeren Modell). Diese „schlaueren“ Reasoning-Modelle werden somit jetzt gerade für komplexere Fragen häufiger zur Anwendung kommen – nach Herstellerangaben stieg die Nutzung dieser stärkeren Modelle unter den zahlenden Usern von 7 % auf 24 %, was insgesamt die Qualität der Ergebnisse steigern sollte. Die neuen Modelle sind wiederum deutlich effizienter geworden, mit stark sinkenden Kosten pro Prompt. Eine Million Token kosteten mit GPT-4 noch 50 Dollar, jetzt nur noch 14 Cent (InvertedStone n. d.; Mollick 2025a, 2025b). Das Modell halluziniert (weiterhin – also Vorsicht! – aber) deutlich seltener als seine Vorgänger: OpenAI gibt hier ca. 1 % Halluzinationen der Antworten statt ca. 5 % bei Vorgängermodellen (o3, 4o) an, je nach Komplexität der Frage und erlaubter „Bedenkzeit“ (OpenAI 2025a, 2025b).

Drittens hat sich die **Internetsuche mit LLMs** deutlich verbessert. Während man früher noch oft über sinnlose oder erfundene Ergebnisse lachte, stellt die Suche von ChatGPT, Google / Gemini oder speziellen Suchanbietern wie Perplexity mittlerweile eine große Zeitersparnis dar: „a useful tool to provide up-to-date answers to questions that are grounded in facts found on the internet, together with the requisite citations — a crucial capability for researchers“ (Korinek 2024, S. 3). Das gilt zunehmend für die stärksten allgemeinen Modelle und erst recht für Anbieter, die auf Forschungsrecherche (und Studierende) spezialisiert sind, wie Elicit oder Paperpal. Auch breite Internet-Recherchen und Textproduktionen sind zunehmend komplett

delegierbar („deep research“), mit deutlichen Auswirkungen auf den Arbeitsprozess in der Wissensarbeit (s. etwa (Schwarcz et al. 2025) für juristische Recherchen, (Korinek 2024) für Ökonomie und (Liang et al. 2025) für PR-Tätigkeiten).

Auch **Studierende nutzen bereits umfangreich Sprachmodelle** für einen breiten Strauß an Zielen (s. *Abbildung 1*). Eine Auswertung der KI-Forscher:innen des Unternehmens Anthropic von einer Million anonymisierter Chats zwischen Usern mit Universitätskonto und dem KI-Bot zeigt typische Nutzungsmuster (Handa, Bent, et al. 2025): Studierende setzen das Sprachmodell **vor allem für anspruchsvolle Tätigkeiten** ein, wie das **Erstellen neuer Inhalte oder das Analysieren komplexer Themen**, was höheren Ebenen der Bloomschen Taxonomie entspricht. Daraus ergibt sich die Herausforderung, sicherzustellen, dass Studierende wesentliche kognitive Aufgaben nicht vollständig an KI delegieren: Aufgaben müssen angepasst und der verantwortungsvolle Umgang mit der Technik eingeübt werden.

 Computer Science	 Natural Sciences & Mathematics
Common Requests	Common Requests
Create and debug C++ programs	Solve and explain statistics problems
Troubleshoot Python code and errors	Work through physics problems with detailed explanations
Teach programming fundamentals with examples	Answer earth science questions
Explain machine learning concepts	Tackle calculus problems with step-by-step explanations
Develop and fix data visualization code	Solve chemistry calculation problems
 Business	 Social Sciences & History
Common Requests	Common Requests
Provide assistance with accounting concepts and problems	Support academic writing about international relations
Analyze business case studies	Explain social science theories
Answer finance questions with calculations	Debug and write Stata code for data analysis
Explain project management concepts	Analyze specific court cases
Create practical negotiation exercises	Solve game theory problems

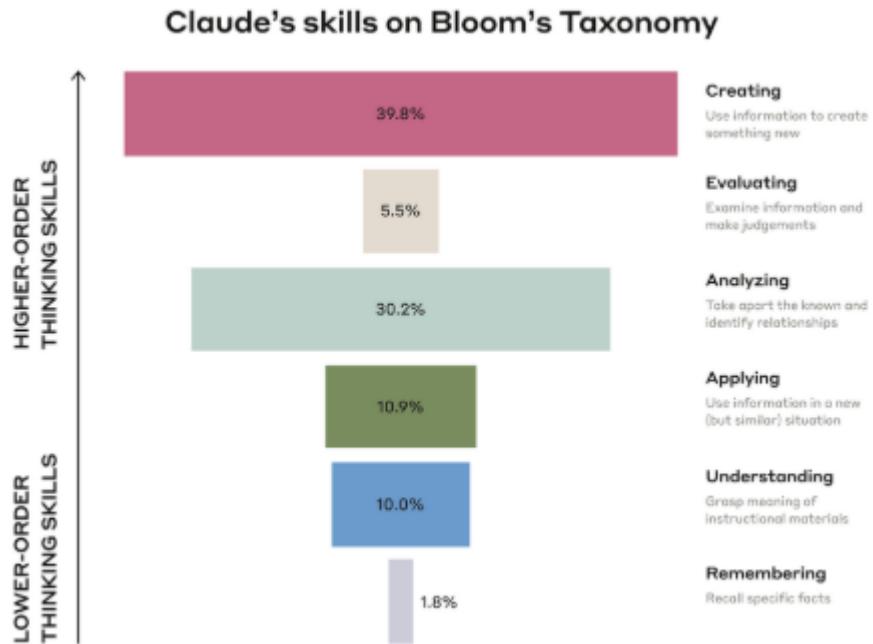
Quelle: Handa, Bent, et al. (2025)]

Auch außerhalb der Hochschule steigt die Nutzung. Eine Reihe von Studien zeigen **erhöhte Produktivität von Büroarbeitenden mit LLM-Unterstützung:** der Kundensupport arbeitet 15% schneller, wenn das Sprachmodell Antwortoptionen vorschlägt und Verweise auf interne technische Dokumentation anbietet (Brynjolfsson, Li, and Raymond 2025), Programmierer programmieren schneller (Peng et al. 2023), Consultants sind produktiver bei komplexen Beratungsprojekten (Dell'Acqua et al. 2023) und Sprachmodelle wie ChatGPT können eine Vielzahl kleiner Aufgaben beschleunigen (Handa, Tamkin, et al. 2025) und werden

insofern gerade zur Texterstellung schon Millionenfach als Hilfsmittel im Beruf genutzt: Von Kundenbewertungen über Pressemitteilungen und Stellenanzeigen (Liang et al. 2025).

Die zunehmende Verwendung von KI in der Lehre hat gute Gründe. Wie oft eine neue Technologie genutzt wird, hängt nach dem **Technology Acceptance Model** (TAM) von der **wahrgenommenen Benutzerfreundlichkeit** (perceived ease of use) und der **wahrgenommenen Nützlichkeit** (perceived usefulness) ab (Marangunić and Granić 2015). Generative KI wie ChatGPT decken sichtlich beide Aspekte ab: Sie sind einfach zu nutzen (Kestin et al. 2024; Lee et al. 2025; Monib, Qazi, and Mahmud 2025; Naddaf 2025b) und erzeugen einen deutlichen Mehrwert, wie Studierende und Lehrende in einer Vielzahl von Umfragen der letzten zwei Jahren berichten (Heidt 2025; Morgan 2024; Ou, Stöhr, and Malmström 2024). Lehrende ziehen nach: Meta-Untersuchungen zeigen ein extremes Wachstum an Publikationen zur Nutzung von LLM im Hochschulalltag (Ma 2025; Ogunleye et al. 2024).

Der Metapher mit dem E-Bike trägt allerdings auch, was die **Risiken und Nebenwirkungen** angeht: Ab wann lässt die maschinelle Unterstützung wichtige Muskeln verkümmern? Solche Gefahren bestehen – wie empirische Studien zeigen, erfordern die neuen Workflows der Wissensarbeit durch KI-Unterstützung auch neue Formen der kritischen Auseinandersetzung mit den Inhalten. Die Analyse von 1 Millionen anonymisierten Studierenden-Chats durch Anthropic (Handa, Bent, et al. 2025) zeigt einerseits, dass **Studierende das LLM vor allem für kognitiv anspruchsvollere Aufgaben einsetzen**, vor allem in den Kategorien „Creating“ und „Analyzing“ (s. Abbildung 2). Dies steht im deutlichen Gegensatz zur Nutzung von einfachen Internetsuchen, die einen Schwerpunkt auf dem Finden einzelner Fakten haben.



Quelle: Handa, Bent, et al. (2025)]

Wie kann man verhindern, dass die Studierenden kritische kognitive Aufgaben allein den KI-Systemen übergeben? Eine Studie von 319 Wissensarbeitern zeigt, dass sich das Gewicht zwischen den **Einzelaufgaben der Wissensarbeit mit LLMsm verschiebt**: Der Aufwand für die Recherchen selbst sinkt, es steigt anderseits der Aufwand für Management-ähnliche Aufgaben: Koordination der Einzelaufgaben für Mensch und Maschine, kritische Prüfung der berichteten Ergebnisse und die Integration von Ergebnissen in den Gesamtprozess (etwa zur Erstellung eines Gesamtberichtes, einer Test-Spezifikation oder eines Protokolls) (Lee et al. 2025). Werden solche neuen Vorgehensweisen nicht geübt, droht ein Rückgang des kritischen Denkens. Lehre heißt in diesem Kontext auch, empfohlene Arbeitsweisen mit der neuen Technik zu üben.

Im Folgenden werden wir zunächst einige **Grundbegriffe** klären: Was sind große Sprachmodelle und was ist mit Begriffen wie Token, Prompt und RAG gemeint? **Welche Modelle** können Lehrende aktuell nutzen und welche **Empfehlungen für Prompts** sind belastbar (Abschnitt 2.5)? Dann fragen wir nach **Zielen**: Welche Art von Wissen und Methoden unterscheidet und empfiehlt die Lernforschung? Welche **didaktischen Wirkmechanismen** können durch KI genutzt werden, um typische Probleme der Hochschullehre anzugehen (Abschnitt 3)? Im Abschnitt 4 schauen wir auf **Praxisbeispiele** für vier Anwendungsfelder von Sprachmodellen an Hochschulen: **KI als Hiwi** (direkte Arbeitserleichterung), **KI als Copilot** (Unterstützung beim Schreiben und Coden) und KI als Tutor (Feedback und Lernunterstützung) sowie **KI als Simulator** (Role Play und Goal Play). Abschließend zeigen wir verschiedene Anwendungen von KI in verschiedenen Kurstypen und gehen auf neue **Herausforderungen für Prüfungen** ein (Abschnitt 5). Im Appendix werden Risiken von KI, rechtliche Rahmenbedingungen und konkrete Prompts und Aufgabenbeispiele aufgeführt.

3 Kapitel 2 – Zentrale Komponenten großer Sprachmodelle

4 Kapitel 3 – Ziele und didaktische Mechanismen - was wollen wir mit KI erreichen?

5 Kapitel 4 – Vier Szenarien: Hiwi, Copilot, Tutor, Simulator

6 Kapitel 5 – Empfehlungen zur Umsetzung

7 Appendix 1 – Risiken und Nebenwirkungen: Halluzinierende Bots, Betrug und Desinformation

8 Appendix 2 – Beispiele für KI-Richtlinien

9 Appendix 3 – Ausgewählte Prompts zur Lehr- und Lernunterstützung

10 Appendix 4 – Detaillierte Beispiele für Aufgabenstellungen mit KI-Unterstützung

- Brown, Peter C., Henry L. Roediger, and Mark A. McDaniel. 2014. *Make it stick: The science of successful learning*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=771951>.
- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. “Generative AI at work.” *The Quarterly Journal of Economics* cjae044.
- Dell’Acqua, Fabrizio, E. McFowland, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, et al. 2023. “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality.” <https://doi.org/10.2139/ssrn.4573321>.
- Gottweis, J., W.-H. Weng, A. Daryin, T. Tu, A. Palepu, P. Sirkovic, et al. 2025. “Towards an AI co-scientist.” *arXiv*. <https://doi.org/10.48550/arXiv.2502.18864>.
- Handa, K., D. Bent, A. Tamkin, M. McCain, E. Durmus, M. Stern, et al. 2025. “Anthropic education report: How university students use claude.” <https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude>.
- Handa, K., A. Tamkin, M. McCain, S. Huang, E. Durmus, S. Heck, et al. 2025. “Which economic tasks are performed with AI? Evidence from millions of claude conversations.” Report. Anthropic. https://assets.anthropic.com/m/2e23255f1e84ca97/original/Economic_Tasks_AI_Paper.pdf.
- Hattie, John. 2023. *Visible learning, the sequel: a synthesis of over 2,100 meta-analyses relating to achievement*. First edition. London New York: Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781003380542>.
- Heidt, A. 2025. “ChatGPT for students: learners find creative new uses for chatbots.” *Nature* 639 (8053): 265–66. <https://doi.org/10.1038/d41586-025-00621-2>.
- Henderson, C., and M. H. Dancy. 2007. “Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics.” *Physical Review Special Topics—Physics Education Research* 3 (2): 020102.
- InvertedStone. n. d. “OpenAI API Pricing Calculator | GPT-5, GPT-5 mini & nano.” <https://invertedstone.com/calculators/openai-pricing>.
- Kestin, G., K. Miller, A. Klales, T. Milbourne, and G. Ponti. 2024. “AI Tutoring Outperforms Active Learning.” <https://doi.org/10.21203/rs.3.rs-4243877/v1>.
- Korinek, Anton. 2024. “LLMs Learn to Collaborate and Reason: December 2024 Update to “Generative AI for Economic Research: Use Cases and Implications for Economists.” *Journal of Economic Literature* 61 (4): 1281–1317. <https://doi.org/10.1257/jel.20231736>.

- Kwa, T., B. West, J. Becker, A. Deng, K. Garcia, M. Hasin, et al. 2025. "Measuring AI Ability to Complete Long Tasks." *arXiv*. <https://doi.org/10.48550/arXiv.2503.14499>.
- Lee, H.-P. H., A. Sarkar, L. Tankelevitch, I. Drosos, S. Rintel, R. Banks, et al. 2025. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers."
- Liang, Weixin, Y. Zhang, M. Codreanu, J. Wang, H. Cao, and J. Zou. 2025. "The Widespread Adoption of Large Language Model-Assisted Writing Across Society." *arXiv Preprint arXiv:2502.09747*. <https://arxiv.org/abs/2502.09747>.
- Ma, T. 2025. "Systematically visualizing ChatGPT used in higher education: Publication trend, disciplinary domains, research themes, adoption and acceptance." *Computers and Education: Artificial Intelligence* 8: 100336. <https://doi.org/10.1016/j.caeari.2024.100336>.
- Marangunić, N., and A. Granić. 2015. "Technology acceptance model: a literature review from 1986 to 2013." *Universal Access in the Information Society* 14 (1): 81–95. <https://doi.org/10.1007/s10209-014-0348-1>.
- McAfee, Andrew. 2024. "Generally Faster: The Economic Impact of Generative AI." Davos Report. Google. <https://ide.mit.edu/wp-content/uploads/2024/04/Davos-Report-Draft-XFN-Copy-01112024-Print-Version.pdf?x76181>.
- Meincke, L., E. R. Mollick, and C. Terwiesch. 2024. "Prompting Diverse Ideas: Increasing AI Idea Variance." Rochester, NY. <https://papers.ssrn.com/abstract=4708466>.
- Meincke, L., E. Mollick, L. Mollick, and D. Shapiro. 2025a. "Prompting Science Report 2: The Decreasing Value of Chain of Thought in Prompting." *arXiv*, 2025a. <https://doi.org/10.48550/arXiv.2506.07142>.
- Mollick, E. 2025a. "GPT-5: It Just Does Stuff." <https://www.oneusefulthing.org/p/gpt-5-it-just-does-stuff>.
- . 2025b. "Mass Intelligence." <https://www.oneusefulthing.org/p/mass-intelligence>.
- Mollick, E., and L. Mollick. 2023. "Assigning AI: Seven Approaches for Students, with Prompts." *arXiv*. <https://doi.org/10.48550/arXiv.2306.10052>.
- Monib, W. K., A. Qazi, and M. M. Mahmud. 2025. "Exploring learners' experiences and perceptions of ChatGPT as a learning tool in higher education." *Education and Information Technologies* 30 (1): 917–39. <https://doi.org/10.1007/s10639-024-13065-4>.
- Morgan, G. 2024. "What Students Want When It Comes To AI - The Digital Education Council Global AI Student Survey 2024." On EdTech Newsletter.
- Naddaf, M. 2025a. "AI is transforming peer review — and many scientists are worried." *Nature* 639 (8056): 852–54. <https://doi.org/10.1038/d41586-025-00894-7>.
- . 2025b. "How are researchers using AI? Survey reveals pros and cons for science." *Nature*, 2025b. <https://doi.org/10.1038/d41586-025-00343-5>.
- Ogunleye, B., K. I. Zakariyyah, O. Ajao, O. Olayinka, and H. Sharma. 2024. "A Systematic Review of Generative AI for Teaching and Learning Practice." *Education Sciences* 14 (6): 636. <https://doi.org/10.3390/educsci14060636>.
- OpenAI. 2025a. "GPT-5 system card." OpenAI. <https://cdn.openai.com/gpt-5-system-card.pdf>.
- . 2025b. "Introducing GPT-5." <https://openai.com/index/introducing-gpt-5/>.
- Ou, A. W., C. Stöhr, and H. Malmström. 2024. "Academic communication with AI-powered language tools in higher education: From a post-humanist perspective." *System* 121: 103225.

<https://doi.org/10.1016/j.system.2024.103225>.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer. 2023. “The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.” *arXiv*. <https://doi.org/10.48550/arXiv.2302.06590>.

Schmidt, B., and R. Tippelt. 2005. “Besser Lehren-Neues von der Hochschuldidaktik?” In *Hochschullandschaft Im Wandel*, 103–14.

Schwarcz, D., S. Manning, P. J. Barry, D. R. Cleveland, J. J. Prescott, and B. Rich. 2025. “AI-Powered Lawyering: AI Reasoning Models, Retrieval Augmented Generation, and the Future of Legal Practice.” <https://doi.org/10.2139/ssrn.5162111>.