

KI-Tutoren in der Lehre: rechtlich sicher, didaktisch wertvoll, technisch innovativ¹

Alexander Gerber², Lars Lorenz³ und Ulrich Pfeiffer⁴

Der unregulierte Einsatz generativer KI im Hochschulalltag wirft erhebliche didaktische, rechtliche, ethische, technische und haushalterische Fragen auf. Als eines der Praxisprojekte von KI:edu.nrw hat sich »Sokratest« diesen Herausforderungen gestellt und eine Art Blaupause entwickelt, wie Lehrende niederschwellig ein kursspezifisches KI-Tutoren-System erstellen können, das technisch innovativ und rechtlich sicher zu betreiben ist. Ausgangspunkt ist dabei immer, welche KI-Assistenz jeweils didaktisch sinnvoll ist. Das System wurde im Sommersemester 2025 in zwei Kursen lernpsychologisch evaluiert. Technisch basiert es auf Open Source Lösungen und dem Prinzip der Retrieval Augmented Generation (RAG). Gewährleistet sind damit Datensouveränität und eine gezielte Beschränkung von Quellen auf validierte Kursinhalte. Bei den Nutzenden zeigte sich u.a. ein signifikanter Anstieg der Fragebereitschaft, während ihre Neugier auf konzeptionell tiefergehende Hintergründe des Erlernten zurückging.

Teaching with AI Tutors: Legally Compliant, Didactically Valuable, Technologically Innovative

The unregulated use of generative AI in higher education raises considerable didactic, legal, ethical, technical and budgetary questions. As one of the practice projects for the regional AI-in-education network in NRW, »Sokratest« has taken on these challenges and developed a sort-of blueprint for how lecturers can create a course-specific AI tutoring system that is technically innovative and legally secure. The fundamental question is always: which AI assistance makes sense didactically? The system was evaluated in terms of learning psychology in two courses in the summer semester of 2025. Technically, it is based on open source solutions and the principle of Retrieval Augmented Generation (RAG). This ensures data sovereignty and a deliberate restriction of sources to validated course content. Users showed, for instance, a

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000-0002-9737-1267

3 ORCID: 0000-0002-0754-3134

4 ORCID: 0009-0006-5381-558X

significant increase in their willingness to ask questions, while their curiosity declined to dive deeper into the conceptual background of course content.

Einleitung: KI-Tutoren im »Wilden Westen« der Hochschullehre

Angesichts der rasanten und weitgehend noch unregulierten Verbreitung generativer Künstlicher Intelligenz (KI) an Hochschulen drängt sich der Eindruck vom »Wilden Westen« auf. Während fast alle Studierenden, die an den Studien von van Garrel und Mayer (2025, 2024) teilgenommen haben, inzwischen in der einen oder anderen Weise auf die großen kommerziellen KI-Modelle als Hilfsmittel zurückgreifen, geschieht dies auch zur Erbringung von Prüfungsleistungen meist noch weitgehend unreguliert, nahezu immer ohne Anleitung oder Aufsicht durch Lehrende und vor allem ohne hinreichende **didaktische Einbettung** (Gerber/Niechoj 2025, 2024). Dabei birgt vor allem der letzte Aspekt erhebliche Risiken: Didaktisch besteht die Gefahr, dass Abkürzungen bei der Erzeugung von Lernprodukten tiefere, reflexive Denkprozesse verhindern und sogenanntes **Cognitive Offloading** vor allem das kritische Denken signifikant negativ beeinflusst (Gerlich 2025), etwa indem erwünschte Denkprozesse komplett ersetzt werden oder sogar didaktisch geradezu entgegengesetzt zu den Lernzielen wirken. Bauer et al. (2025) beschreiben dies als »Inversion Effect« – die kritischste der vier möglichen Auswirkungen von KI in Lernprozessen gemäß des »ISAR«-Modells, das wiederum auf früheren etablierten Modellen der Didaktikforschung beruht (u. a. ICAP und SAMR, cf. Chi/Wylie 2014; Sailer et al. 2024). Neben solchen modellhaften Betrachtungen liegen mittlerweile auch alarmierende empirische Befunde vor, dass unreflektierte KI-Nutzung dem Lernprozess nachhaltig schaden kann. So sprechen Fan et al. (2024) nach ihren randomisierten Experimenten von der Gefahr einer KI-induzierten »metakognitiven Faulheit«. Stadler et al. (2024) schlussfolgern aus ihren Experimenten sogar, dass die Fähigkeit von Lernenden zum Treffen komplexer Entscheidungen nach Nutzung generativer KI geringer ist als jene von Lernenden, die mit einer klassischen Suchmaschine recherchieren. Es kann daher als breiter wissenschaftlicher Konsens bezeichnet werden, dass eine unzureichend angeleitete und nicht an didaktischen Zielen ausgerichtete Nutzung von KI an der Hochschule den Lernerfolg von Studierenden nachhaltig gefährdet.

Auch rechtlich ergeben sich im Kontext der Lehre vielfältige Bedenken u. a. hinsichtlich des **Urheberrechts** (beim Hochladen von Kursmaterialien in kommerzielle Dienste), des **Datenschutzes** (etwa durch die Verarbeitung von Nutzungsinformationen außerhalb der EU) oder des **Prüfungsrechts** (etwa aufgrund unklarer und/oder uneinheitlicher Regelungen). Auch technische und finanzielle **Abhängigkeiten** von kommerziellen Anbietern sind zu bedenken, da deren Lösungen in der Praxis nahezu nie anpassbar und vergleichsweise intransparent sind.

Um diesen Herausforderungen gezielt zu begegnen, wurde im Rahmen des Projekts Sokrates[†] ein Prozessmodell entwickelt, das es Lehrenden ermöglicht, mit wenigen Klicks und auch ohne technische Vorkenntnisse einen eigenen KI-Tutor zu erstellen, der auf ihren jeweiligen Kurs zugeschnitten ist und dabei die oben erwähnten rechtlichen, ethischen und didaktischen Aspekte von vornherein mit berücksichtigt. Der Ansatz von Sokrates[†] vereint vier zentrale Dimensionen:

1. Klarer didaktischer Mehrwert (in dem als erstes evaluierten Prototypen demonstriert durch die Förderung kritischen Denkens mittels sokratischer Dialoge);
2. Technische Innovation durch einen niederschweligen und nahezu beliebig skalierbaren Open-Source-Ansatz;
3. Rechtssicherheit durch DSGVO-konformen Datenschutz und urheberrechtliche Sicherungsmechanismen;
4. Förderung von Ethik und KI-Kompetenzerwerb durch co-kreative Entwicklung, Nutzung und Reflexion durch Studierende.

Technische Umsetzung: Ein Open-Source-Ansatz für die Datensouveränität

Das technische Herzstück von Sokrates[†] ist die **Retrieval-Augmented Generation (RAG)**. Dieses Verfahren lässt sich am besten als eine Art »Open-Book-Klausur« für die KI beschreiben. Es adressiert direkt das Problem der mangelnden Verlässlichkeit und reduziert die Neigung von Sprachmodellen zu halluzinieren (Shuster et al. 2021). Anstatt auf sein allgemeines und oft unzuverlässiges oder wenig spezifisches Weltwissen zuzugreifen, generiert das Sprachmodell Antworten weitestgehend auf Basis einer vorab definierten, validierten Wissensbasis. Diese besteht bei Sokrates[†] aus den von Lehrenden hochgeladenen Kursmaterialien (Vorlesungs-Skripte, wissenschaftliche Artikel, Lehrbücher, Foliensätze etc.). Jede von der KI generierte Antwort kann dadurch mit direkten Zitaten und Verweisen auf die im Kurs offiziell genutzten Quellen belegt werden, was maximale Transparenz und Nachvollziehbarkeit gewährleistet.

Die technische Pipeline gliedert sich in **drei Schritte**. Zuerst erfolgt die **Vorverarbeitung der Inhalte**: Lehrende laden ihre Dokumente hoch, die das System automatisiert in logische, semantische Abschnitte – sogenannte Chunks – zerlegt. Im zweiten Schritt, der **Einbettung und Speicherung**, wird jeder Chunk in eine numerische Repräsentation (Vector) umgewandelt und in einer dafür optimierten Datenbank gespeichert. Der dritte Schritt umfasst die **Anfrage und Antwort**: Stellen Studierende eine Frage im Chat, findet das System die relevantesten Chunks und übergibt diese an die KI, das Large Language Model (LLM). Das Sprachmodell for-

muliert daraufhin eine Antwort, die auf diesen Inhalten basiert. Die Quellen können nun präzise angegeben werden, da im Prozess bekannt ist, welche Dokumente und einzelnen Chunks an das LLM übergeben wurden. Technisch werden die Chunks als Teil des System Prompts übergeben, welcher damit direkt den Kontext und das Verhalten der KI steuert. Für Sokrates[†] wurde dieser Prompt unter Verwendung eines literaturbasierten, didaktisch fundierten Ansatzes entwickelt und implementiert ein strukturiertes sokratisches Fragemodell, das sich an der weiterentwickelten Bloom'schen Taxonomie orientiert (Anderso/Krathwohl 2001). Lernende werden so vom grundlegenden Verständnis zur kritischen Bewertung und kreativen Synthese geführt. Der System Prompt wurde durch ein strukturiertes Testverfahren gestaltet, bestehend aus 50 simulierten Interaktionen in Blöcken von jeweils fünf Dialogen, und iterativ verfeinert.

Um die volle Datensouveränität und rechtliche Konformität zu sichern, basiert das gesamte System auf gängigen **Open-Source-Technologien** (u.a. React, Qdrant, Postgres, Docker) und wird auf einem Server in Deutschland **selbst gehostet**. Für die Dialogführung werden Open-Source-Sprachmodelle genutzt, die über eine sichere Schnittstelle der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) angebunden sind. Dadurch wird sichergestellt, dass keine Anfragen oder Daten an kommerzielle Anbieter gesendet werden. Eine integrierte Nutzer- und Dokumentenverwaltung sichert das System und alle darin vektorisierten Inhalte urheberrechtlich ab und beschränkt den Zugang pro Kurs auf die eingeschriebenen Teilnehmenden.

Lernpsychologische Begleitforschung: Methodik und vorläufige Ergebnisse

Entwicklung und Einsatz von Sokrates[†] im Sommersemester 2025 wurden lernpsychologisch evaluiert. Das Forschungsdesign umfasste eine Pre-Post-**Interventionsstudie** mit 62 teilnehmenden Studierenden aus zwei Pilot-Kursen (Bachelor und Master). Mittels quantitativer Befragungen wurden Lernpräferenzen, KI-Einstellungen und Wissen vor und nach der Nutzung des Tutors erfasst. Ergänzend wurden zwecks qualitativer Triangulation zwei **Fokusgruppen** durchgeführt. Außerdem reflektierten die Studierenden auf insgesamt ca. 400 Seiten ihre Nutzungserwartungen und -erfahrungen in Bezug auf die KI in Form individueller **Lerntagebücher**. Besonders aufschlussreich sind hier die Gegenüberstellungen der Verwendung von Sokrates[†] mit frei zugänglichen (und in der Regel kommerziellen) Sprachmodellen. Die inhaltsanalytische Auswertung der Tagebücher ist zum Zeitpunkt der Erstellung des vorliegenden Beitrags ebenso in Arbeit wie die Auswertung der insgesamt 883 von den Studierenden verfassten Prompts und der dazugehörigen Antworten. Die zu erwartenden Einblicke in die Interaktionsdynamik zwischen Studieren-

den und KI sollen in der weiteren Entwicklung das didaktische Verhalten des KI-Tutors unterstützen. Da es sich nicht um ein Laborexperiment handelte, sondern um zwei regulär angebotene Kurse, haben sich zwischen Pre- und Post-Messung etliche Studierende (vor allem im Bachelor-Modul) entschieden, den Kurs nicht abzuschließen. Dies ist an unserer Hochschule in den ersten Semesterwochen üblich und erklärt die vergleichsweise hohe Abbruchrate zwischen den beiden Messzeitpunkten.

Schlüsselergebnis 1: Deutlich erhöhte Fragebereitschaft

Eines der zentralen Ziele des Projekts war es, eine sichere Lernumgebung zu schaffen, in der Studierende ohne Hemmungen Fragen stellen können. Dieses Ziel wurde eindrucksvoll erreicht. Die Bereitschaft, dem KI-Tutor Fragen zum Kursinhalt zu stellen, war deutlich höher, als sich mit diesen Fragen an Lehrende zu wenden: Der Mittelwert auf einer 7-stufigen Skala erhöhte sich von 5,15 (pre) auf 6,21 (post), was einer signifikanten Effektstärke von **Cliff's Delta = 0,43** entspricht. In den Fokusgruppen bestätigten Studierende, dass der KI-Tutor selbst bei subjektiv als »dumm« empfundenen Fragen unterstützend reagierte, was die Wahrnehmung des eigenen Frageverhaltens positiv beeinflusste. Erste Auswertungen der Lerntagebücher unterstreichen diese Lernerfahrung eines geschützten und vertrauenserweckenden Raums. Allerdings setzt diese Erfahrung voraus, das zumindest grundlegende Kenntnisse vorhanden sind, um überhaupt in der Lage zu sein, eine relevante Frage formulieren zu können.

Schlüsselergebnis 2: KI-Kurzschlüsse und die »Antwortmaschine«

Gleichzeitig offenbarte die Studie eine bedenkliche Kehrseite. Die Neugier der Studierenden, das »Warum« hinter einem Konzept zu verstehen, anstatt nur das »Was« zu lernen, sank über das Semester hin deutlich. Der Mittelwert fiel von 6,15 (pre) auf 4,93 (post), was einer signifikanten negativen Effektstärke von **Cliff's Delta = -0,57** entspricht. Dieses Ergebnis deutet darauf hin, dass der Tutor trotz seines explizit sokratischen Designs von einigen Studierenden als zu effiziente Antwortmaschine wahrgenommen wurde, was wiederum die intrinsische Motivation zur tieferen, eigenständigen Auseinandersetzung zu reduzieren scheint und somit die feine Gratwanderung zwischen hilfreicher Unterstützung und der Förderung intellektueller Eigenständigkeit verdeutlicht. Sprachmodelle wie Llama 3.3 (70B) ließen sich hier leichter überzeugen, ausführliche Antworten zu geben, als neuere LLMs, die auch »Reasoning« unterstützen. In keinem einzigen Fall trafen diese Modelle gänzlich falsche Aussagen; sie stellten aber in einzelnen Fällen etwas zu viele Informationen bereit.

Schlüsselergebnis 3: Gestärktes Vertrauen und KI-Kompetenz

Die Nutzung von Sokrates^f über das Semester hinweg stärkte die Selbstwahrnehmung und Akzeptanz von KI-Werkzeugen signifikant. Studierende fühlten sich qualifizierter, KI beim Lernen einzusetzen ($d=0,45$), und sahen mehr Vorteile in KI-Systemen ($d=0,37$). Zwar zeigen die Ergebnisse eine positive Tendenz im selbst wahrgenommenen KI-Kompetenzerwerb, unterstreichen aber auch die Verantwortung, die mit dem Einsatz solcher Systeme einhergeht. Die Selbstwahrnehmung und das Nutzungsverhalten im Bereich von KI-Werkzeugen bei Studierenden werden direkt beeinflusst, was unbedingt unter Anleitung und verantwortungsvoll geschehen muss. Auch hier zeigen erste Auswertungen der Lerntagebücher die von nahezu allen Studierenden beschriebene wahrgenommene Kompetenzerweiterung hinsichtlich Prompt Engineering und kritischer Analyse von KI-generierten Inhalten.

Diskussion und Ausblick: Lessons Learned für die »Bauanleitung«

Die bisherigen Ergebnisse zeichnen ein differenziertes Bild. Die Stärken von Sokrates^f liegen darin, dass ein **sicherer Raum** für Fragen geschaffen und eine rechtlich und technisch **souveräne Lösung** bereitgestellt wird, die von den Studierenden als **vertrauenswürdig** wahrgenommen wird und die ihre selbstwahrgenommene **KI-Kompetenz** signifikant steigert. Eine der größten Herausforderungen liegt in der didaktischen Feinjustierung: Die Balance zwischen direkter Hilfe und sokratischer Führung ist weiter zu optimieren, um einen Verlust tiefergehender Neugier zu vermeiden. Auch das Studierendenfeedback lieferte viele wertvolle Ansatzpunkte zur Weiterentwicklung, etwa verschiedene Modi (einen für direkte Antworten und einen für kritisches Hinterfragen). Ein weiterer Aspekt ist die Rolle von **Learning Analytics (LA)**. Durch die Analyse anonymisierter Dialogmuster können Lehrende erkennen, welche Themen den Studierenden besondere Schwierigkeiten bereiten. Ziel ist hierbei kein Überwachungsinstrument, sondern ein dynamisches, anonymisiertes LA-System, das Lehrenden direktes Feedback zur Verbesserung ihrer Lehre gibt (z.B. aggregierte Erkenntnisse wie: »Fragen zu Themenaspekt X wurden insgesamt 22-mal so oft gestellt, wie Fragen zu Themenaspekt Y. Darauf sollte in der Vorlesung ggf. nochmals eingegangen werden.«). Die nächsten Schritte umfassen die Verfeinerung des Prozessmodells zu einer finalen »Bauanleitung«, die es Lehrenden ermöglicht, ohne externe Hilfe Tutoren zu erstellen – mittelfristig auch über den prototypisch bereits evaluierten sokratischen Ansatz hinaus hinsichtlich anderer didaktischer Einsatzszenarien. Technisch ist die Skalierung für weitere Kurse und Hochschulen bereits jetzt möglich. Wir laden

interessierte Kolleginnen und Kollegen daher herzlich zur weiteren Kooperation ein.

Literatur

- Anderson, Lorin W./Krahtwohl, David R. (2001): *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Longman.
- Chi, Michelene/Wylie, Ruth (2014): »The ICAP framework: Linking cognitive engagement to active learning outcomes«, in: *Educational Psychologist*, 49(4), S. 219–243.
- Fan, Yizouh/Tang, Luzhen/Le, Huixiao/Shen, Kejie/Tan, Shufang/Zhao, Yueying/Shen, Yuan/Li, Xinyu/Gašević, Dragan (2024): »Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance«, in: *British Journal of Educational Technology*, 00, S. 1–42.
- Gerber, Alexander/Niechoj, Torsten (2024): »KI auf dem Campus: Chancen und Herausforderungen«. in: *Die Neue Hochschule*, S. 30–33. https://www.hlb.de/fileadmin/hlb-global/downloads/dnh/full/2024/DNH_2024-3.pdf
- Gerber, Alexander/Niechoj, Torsten (2025): »KI auf dem Campus: Von der Empirie zur Strategie«. in: Robert Queckenberg/Jonas Leschke/Malte Persike (Hg.), *Learning Analytics, Artificial Intelligence und Data Mining in der Hochschulbildung*. Beiträge zur Learning AID 2024, Bielefeld: transcript Verlag.
- Gerlich, Michael (2025): »AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking«, in: *Societies*, 15(1), 6.
- Sailer, Michael/Maier, Rebecca/Berger, Sonja/Kastorff, Tamara/Stegmann, Karsten (2024): »Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education«, in: *Learning and Individual Differences*, 112(102446).
- Shuster, Kurt/Poff, Spencer/Chen, Moya/Kiela, Douwe/Weston, Jason (2021): »Retrieval Augmentation Reduces Hallucination in Conversation«, in: *Conference on Empirical Methods in Natural Language Processing*, arXiv.
- Stadler, Matthias/Bannert, Maria/Sailer, Michael (2024): »Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry«. in: *Computers in Human Behavior*, 108386.
- von Garrel, Jörg & Mayer, Jana (2024): »Which features of AI-based tools are important for students? A choice-based conjoint analysis«, in: *Computers and Education: Artificial Intelligence*, 7(100311).

von Garrel, Jörg/Mayer, Jana (2025): Künstliche Intelligenz im Studium – Eine quantitative Längsschnittstudie zur Nutzung KI-basierter Tools durch Studierende, Darmstadt: Hochschule Darmstadt. Preprint.