Jonas Leschke,
Robert Queckenberg, Malte Persike (Hg.)

LEARNING ANALYTICS, ARTIFICIAL INTELLIGENCE UND DATA MINING IN DER HOCHSCHULBILDUNG

Beiträge zur Learning AID 2025

[transcript] Zukunft der Hochschule

Jonas Leschke, Robert Queckenberg, Malte Persike (Hg.)
Learning Analytics, Artificial Intelligence und Data Mining in der Hochschulbildung

Editorial

Die Hochschule befindet sich im Wandel: Studien- und Verwaltungsformen im Sinne des New Public Managements, die Digitalisierung sowie Forderungen nach mehr Diversität, Inklusion und Nachhaltigkeit stellen Forschung und Lehre vor große Herausforderungen.
Die Reihe **Zukunft der Hochschule** legt den Fokus auf die Risiken und Chancen dieser Entwicklungen und fragt nach der Zukunft unseres Hochschulsystems. Neben kritischen Perspektiven auf die neoliberalen Umstrukturierungsprozesse bietet sie Publikationen ein Forum, die Szenarien für eine sozial-ökologische Transformation der Hochschule entwerfen und nach inklusiveren Bildungszugängen und -formaten fragen. Zugleich ist dies der Ort in unserem Programm, an dem die Hochschulen mit außeruniversitären Forschungs- und Bildungseinrichtungen sowie zivilgesellschaftlichen Akteur:innen in Kontakt treten, um Strategien für einen partizipativen Wissenstransfer zu entwickeln.

Jonas Leschke ist Leiter des Zentrums für Lehren und Lernen (ZLL) an der HAW Hamburg. Er verantwortet und koordiniert die Handlungsfelder des ZLL hochschuldidaktische Weiterbildung und Lehrcoaching, digital gestütztes Lehren, Lernen und Prüfen sowie Lehrinnovation und Curriculumsentwicklung.
Robert Queckenberg koordiniert am Zentrum für Wissenschaftsdidaktik der Ruhr-Universität Bochum das vom Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen geförderte Konsortialprojekt KI:edu.nrw, das sich mit den Themen Learning Analytics und Künstliche Intelligenz in der Hochschulbildung beschäftigt und die Learning-AID-Tagung ausrichtet.
Malte Persike ist wissenschaftlicher Leiter am Center für Lehr- und Lernservices (CLS) der RWTH Aachen University. Er verantwortet und koordiniert das Forschungsprogramm des CLS und beschäftigt sich insbesondere mit der datengestützten Optimierung des Lehrens, Lernens und Prüfens.

Jonas Leschke, Robert Queckenberg, Malte Persike (Hg.)

Learning Analytics, Artificial Intelligence und Data Mining in der Hochschulbildung

Beiträge zur Learning AID 2025

[transcript]

Gefördert durch:

Ministerium für
Kultur und Wissenschaft
des Landes Nordrhein-Westfalen

Ein Kooperationsvorhaben empfohlen durch die:

DIGITALE HOCHSCHULE NRW

INNOVATION DURCH KOOPERATION

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://dnb.dnb.de/ abrufbar.

Dieses Werk ist unter der Creative-Commons-Lizenz BY 4.0 lizenziert. Für die ausformulierten Lizenzbedingungen besuchen Sie bitte die URL https://creativecommons.org/licenses/by/4.0/.

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z.B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

2026 © Jonas Leschke, Robert Queckenberg, Malte Persike (Hg.)

transcript Verlag | Hermannstraße 26 | D-33602 Bielefeld | live@transcript-verlag.de

Umschlagkonzept: Maria Arndt
Korrektorat: Jana Prokop
Druck: Elanders Waiblingen GmbH, Waiblingen
https://doi.org/10.14361/9783839458761
Print-ISBN: 978-3-8376-8045-4 | PDF-ISBN: 978-3-8394-5876-1
Buchreihen-ISSN: 2943-4882 | Buchreihen-eISSN: 2943-4890

Gedruckt auf alterungsbeständigem Papier mit chlorfrei gebleichtem Zellstoff.

Inhalt

Editorial
Jonas Leschke, Robert Queckenberg und Malte Persike 9

Keynote

GenKI und Learning Analytics zusammendenken
Ein konzeptioneller Impuls für interaktionszentrierte Learning Analytics
Sabine Seufert 15

Künstliche Intelligenz

Von Betroffenen zu Beteiligten
Ein studentisch entwickeltes Dialogformat zu KI in der Hochschullehre
Sarah Becker und Katharina Westphal 31

Die KI-Experimentierumgebung an der FernUniversität in Hagen
Erste Erfahrungen aus der Lehre mit generativer KI
Anna Hinzmann und Ferdal Özcelik 41

Implementierung, Nutzung und Herausforderungen eines hochschuleigenen KI-Interfaces
Isabella Buck, Nina Christ, Alexander Kaib und Lisa Ulzheimer 49

KI-Tutoren in der Lehre: rechtlich sicher, didaktisch wertvoll, technisch innovativ
Alexander Gerber, Lars Lorenz und Ulrich Pfeiffer 57

Offene Künstliche Intelligenz als Chance für Bildungsprozesse
Beatrix Busse, Ingo Kleiber und Kathrin Andree 65

Das Sprechen über KI und AI Literacy
Didaktische Konsequenzen aus einer Korpusanalyse
Sophia Hercher 71

Förderung von AI Literacy bei Studierenden
Ein Modell für die Integration von KI-Kompetenzen in die Hochschulbildung
Miriam Stehling, Claudia Berg, Manuel Geisler, Miriam Hägerbäumer, Nils Aschhoff und Vera Lenz-Kesekamp 79

Lehr-/Lernangebote zum KI-gestützten Erwerb Akademischer Schreibkompetenz in Fremdsprachen
Torsten Kirchner und Géraldine Stich-Desmarchelier 89

Fostering AI Literacy in Higher Education
The Role of Students' Usage Patterns
Sittipan Yotyodying, Maria Zirenko, Sabine Fabriz und Holger Horz 97

KIViPro – KI-unterstützte Videoproduktion fördern und lernen
Hendrik Bunke, Margret Plank, Jens Kösters, Matti Stöhr und Zoë Kittelmann 111

Singularity – Ein Serious Game für AI Literacy
Laura Platte und Malte Persike 119

Aus der Rolle gefallen?
Entwicklung von KI-gesteuerten Spielfiguren mit Studierenden
Luke West und Esther Fink 127

Analyse qualitativer Daten mit KI-Unterstützung
Evolution statt Revolution qualitativer Evaluationsmethoden in der Qualitätssicherung der Philipps-Universität Marburg
Florian Engel und Tristan Dohnt 137

KI in der Lehre – Eine aktuelle Übersicht über die Nutzung von KI im Studienkontext
Rischert Enes, Arnim Spengler, Felix Meckmann, Thanh Viven Le-Vu und Khaoula Benmaarouf 145

Systematische Evaluation prompt-basierter, durch KI generierte Multiple-Choice-Fragen in der Medizinischen Biochemie
Masoud Darabi und Shabnam Fayezi 153

Learning Analytics

Integration von Learning Analytics in der Hochschullehre mittels POLARIS
Erfahrungen aus dem Praxisprojekt ISLAnD_BWL
Holger Ketteniß und Johanna Marie Kunsmann 163

AnnoPy – Potenziale von KI-basierter Learning Analytics für die Förderung wissenschaftlicher Textkompetenz
Oliver Scholle, Sara Rezat, Sebastian Rezat und Dominik Niehus 171

Wie Learning Analytics und KI selbstreguliertes Lernen transformieren
Didaktische Potenziale videobasierter Lernumgebungen in der Lehrkräftebildung
Manuel Oellers, Robin Junker, Jennifer Janeczko, Christina Dückers und Manfred Holodynski 183

Wie und von wem werden digitale Unterstützungsangebote zur Testung und Auffrischung schulischer Kompetenzen in der Hochschule genutzt?
Eine Evaluation anhand von Klickdaten und Befragungen
Carolin Horsthemke, Elena Schmitt, Jakob Schwerter und Hanna Gaspard 197

Editorial

Jonas Leschke[1], Robert Queckenberg[2] und Malte Persike[3]

Die vierte Learning AID-Tagung

Am zweiten und dritten September 2025 fand die vierte Learning AID-Tagung an der Ruhr-Universität Bochum statt. Die Learning AID ist die mit über 350 Teilnehmenden größte deutschsprachige Fachtagung mit dem thematischen Fokus auf *Learning Analytics, Artificial Intelligence und Data Mining in der Hochschulbildung* und wird seit ihrer ersten Ausführung durch das vom Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen geförderte Projekt KI:edu.nrw (Salden/Leschke/Persike 2024) ausgerichtet. Die Ziele der Tagung sind der interdisziplinäre Praxis- und Forschungsdialog über Künstliche Intelligenz und Learning Analytics in der Hochschullehre sowie die fach- und community-übergreifende Vernetzung der Hochschulakteur*innen, die diese Themenfelder prägen (Leschke/Queckenberg/Persike 2025).

Um diese Ziele erreichen zu können, müssen in der Ansprache möglichst viele unterschiedliche Netzwerke zur Teilnahme an der Tagung erreicht werden. Zu diesem Zweck fand im Vorfeld an die vierte Ausgabe der Tagung am ersten September zusätzlich zu den beiden Tagungstagen ein Community-Tag statt. Sieben Netzwerke und Communities organisierten ebenfalls an der Ruhr-Universität Bochum für ihre Mitglieder und interessierte Gäste Austauschformate zu den Tagungsthemen. Vertreten waren die *Deutsche Gesellschaft für Hochschuldidaktik (dghd), die Gesellschaft für Schreibdidaktik und Schreibforschung (gefsus), die Gesellschaft für Information, Beratung und Therapie an Hochschulen e. V. (GIBeT), die Gesellschaft für Hochschulforschung (GfHf), die Sektion Medienpädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE), die DigitalChangeMaker des Hochschulforums Digitalisierung sowie die Rechtsinformationsstelle ORCA.nrw.* Am Community-Tag nahmen bereits rund 100 Teilnehmende teil und ein großer Teil dieser Personen besuchte die anschließende Learning AID-Tagung.

1 ORCID: 0000–0002-6343-117X
2 ORCID: 0009–0006-5171-438X
3 ORCID: 0000–0002-7825-089X

Die Tagung selbst zog auch in diesem Jahr wieder ein großes Interesse auf sich, was sich bereits beim Call for Submissions zeigte. Für die Learning AID 2025 wurden 118 Einreichungen für Impulsvorträge, Workshops, Poster und Demonstrationen übermittelt. Auch in diesem Jahr erfolgte die Auswahl der Impulsvorträge und Workshops durch ein double-blind Peer-Review Verfahren durch die Projektmitarbeitenden von KI:edu.nrw, Mitarbeitenden des Zentrums für Wissenschaftsdidaktik der Ruhr-Universität Bochum und den Einreichenden. Die Poster und Demonstrationen wurden von den Tagungsorganisator*innen begutachtet und sind daher im vorliegenden Tagungsband nicht berücksichtigt worden. Insgesamt wurden 71 Einreichungen im Tagungsprogramm aufgenommen. Der vorliegende Tagungsband enthält 20 ausgearbeitete und von den Herausgebenden begutachtete Publikationen. Diese zeichnen die leitenden Themen im Diskurs um Künstliche Intelligenz und Learning Analytics in der Hochschulbildung in den Jahren 2024 und 2025 nach: Förderung von KI-Kompetenz, der kombinierte Einsatz von Learning Analytics und Künstlicher Intelligenz sowie die studentische Partizipation im Umgang mit generativer KI. Diese leitenden Themen stellten zudem den inhaltlichen Fokus der drei rahmenden Panelsessions dar; eine Session zur aktuellen wissenschaftspraktischen Perspektive von KI in der Hochschulbildung, eine Session zur politischen Perspektive von KI und Learning Analytics in der Hochschulbildung und eine interaktive Session mit Expert*innen und Einbindung des Publikums zu »Future Quests« der Hochschullehre.

Nachfolgend erhalten Sie einen Überblick darüber, was Sie im Tagungsband zur Learning AID 2025 erwartet. Ein besonderer Dank gilt Maja Fietzek für ihre Unterstützung bei der Endredaktion der Manuskripte.

Ausblick in den Tagungsband

Keynote

Seufert spannt im Rahmen ihres Beitrags zu ihrer Keynote das Forschungs- und Entwicklungsfeld der Kombination von Learning Analytics und generativer KI auf. Dabei arbeitet sie heraus, welches Potenzial im AI-Integrated Learning Analytics, in der systematischen Verschränkung von Learning Analytics und KI zum lernwirksamen Einsatz der Technologien, liegt.

Künstliche Intelligenz

Becker und Westphal stellen in ihrem Beitrag ein Workshopkonzept vor, um unterschiedliche Statusgruppen an Hochschulen, von Studierenden bis hin zu strategischen Entscheider*innen, in einen strukturierten Dialog zum Umgang mit KI an

Hochschulen zu bringen. Zudem diskutieren sie die Ergebnisse der Erprobung des Konzepts im Rahmen der Learning AID-Tagung.

Hinzmann und Özcelik präsentieren Ergebnisse aus der Begleitforschung der Einführung einer an der Fernuniversität Hagen entwickelten KI-Experimentierumgebung. Ein besonderer Fokus dieser Experimentierumgebung liegt auf dem lokalen Betrieb von Large Language Models und der didaktisch gerahmten Integration in Moodle.

Buck, Crist, Kaib und Ulzheimer diskutieren in ihrem Beitrag Evaluationsergebnisse aus der Nutzung eines hochschuleigenen KI-Interfaces und zeigen dabei insbesondere Herausforderungen für die hochschulweite Akzeptanz auf.

Gerber, Lorenz und Pfeiffer stellen einen sokratischen KI-Tutor vor, der durch Retrieval Augmented Generation veranstaltungsspezifische Informationen in seinen Ausgaben berücksichtigt und anstatt Lösungen lernunterstützende Texte und Anregungen generiert. In der Nutzungsevaluation konnte unter anderem festgestellt werden, dass Lernende dazu angeregt werden, Fragen zu stellen.

Busse, Kleiber und Andree arbeiten in ihrem Beitrag anhand von drei Bezugspunkten das Potenzial von offener KI für Bildungsprozesse heraus und zeigen auf, dass der Wert von offener KI für Hochschulen und Bildungsprozessen in mehr als dem technischen Mehrwert liegt.

Hercher untersucht in ihrem Beitrag korpuslinguistisch, wie über generative KI auch in Bildungssituationen gesprochen wird und leitet didaktische und sprachliche Konsequenzen für die Gestaltung von Lernaufgaben ab.

Stehling, Berg, Geisler, Hägerbäumer, Aschhoff und Lenz-Kesekamp präsentieren die Entwicklung und das Ergebnis eines KI-Kompetenzmodells. Dabei betonen sie, dass Hochschulen in Abhängigkeit der Breite ihres Fächerspektrums und der Heterogenität der Studierendenschaft individuelle Lösungen entwickeln müssen.

Kirchner und Stich-Desmarchelier stellen einen Ansatz zur lernwirksamen Integration von KI-Tools zum Erwerb akademischer Schreibkompetenz im Fremdsprachenunterricht durch den Einsatz eines E-Portfolios vor.

Yotyodying, Zirenko, Fabriz und Horz untersuchen empirisch, ob die Nutzung von KI-Tools mit der Entwicklung von KI-Kompetenz einhergeht und kommen zu dem Schluss, dass weniger die Nutzungshäufigkeit, als die Nutzungsqualität ausschlaggebend für die Kompetenzentwicklung ist.

Bunke, Plank, Kösters, Stöhr und Kittelmann beschreiben in ihrem Beitrag ein als OER veröffentlichtes Lernangebot zur Produktion von Lehr- und Lernvideos mittels generativer KI und präsentieren die Evaluationsergebnisse von 100 Lehrenden, die das Angebot genutzt haben.

Platte und Persike stellen mit Singularity ein Serious Game vor, das Studierende bei der Entwicklung von AI Literacy unterstützt und fachübergreifend einsetzbar ist.

West und Fink präsentieren in ihrem Beitrag einen Ansatz, wie Studierende KI-gesteuerte Spielfiguren entwickeln können und beschreiben das Lernpotenzial dieses Ansatzes.

Engel und Dohnt analysieren qualitative Daten mittels generativer KI und reflektieren diesen Ansatz für die Qualitätssicherung an Hochschulen.

Enes, Spengler, Meckmann, Le-Vu und Benmaarouf präsentieren Befragungsergebnisse einer deutschlandweiten Befragung über den Einsatz von generativer KI mit über 1.000 Teilnehmenden an Hochschulen und leiten auf Grundlage dieser Ergebnisse ein praxisorientiertes Konzept für Hochschulen ab.

Darabi und Fayezi vergleichen KI-generierte Multiple-Choice-Fragen unterschiedlicher Chatbots im medizinisch-biochemischen fachlichen Kontext und zeigen Stärken und Grenzen der unterschiedlichen Systeme auf.

Learning Analytics

Kettenieß und Kunsmann beschreiben die Integration vom Learning Analytics-System POLARIS zur individuellen Empfehlung zur weiteren Bearbeitung von Lernmaterialien und somit skalierbaren individuellen Lernangeboten.

Scholle, Rezat, Rezat und Niehus stellen mit AnnoPy ein digitales Tool zum kollaborativen Lesen und Analysieren von Texten vor und vergleichen anhand von Nutzungsdaten die händischen Analysen mit den AnnoPy-gestützten Analysen.

Oellers, Junker, Janeczko, Dückers und Holodynski beschreiben in ihrem Beitrag, wie Learning Analytics das Zusammenspiel synchroner und asynchroner Lernphasen durch formatives Feedback in der Lehrkräftebildung unterstützen kann.

Horsthemke, Schmitt, Schwerter und Gaspard untersuchen anhand von Klickdaten und Befragungen die Nutzung von digitalen Unterstützungsangeboten zur Auffrischung von schulischen Kompetenzen in der Studieneingangsphase.

Literatur

Leschke, Jonas/Queckenberg, Robert/Persike, Malte (2025): Editorial: Eine Fachtagung für Learning Analytics und Künstliche Intelligenz in der Hochschulbildung, in: Robert Queckenberg/Jonas Leschke/Malte Persike (Hg.), Learning Analytics, Artificial Intelligence und Data Mining in der Hochschulbildung – Beiträge zur Learning AID 2024, transcript Verlag, S. 9–16.

Salden, Peter/Leschke, Jonas/Persike, Malte (2024): Das Projekt KI:edu.nrw – Rückblick für einen Ausblick, in: Peter Salden/Jonas Leschke (Hg.), Learning Analytics und Künstliche Intelligenz in Studium und Lehre, Springer Fachmedien Wiesbaden, S. 3–24.

Keynote

GenKI und Learning Analytics zusammendenken

Ein konzeptioneller Impuls für interaktionszentrierte Learning Analytics[1]

Sabine Seufert

Der zunehmende Einsatz generativer Künstlicher Intelligenz (GenKI) in Schule und Hochschule stellt etablierte Prüfungs- und Feedbackpraktiken infrage. Der Fokus aktueller Debatten liegt vor allem auf der Absicherung bestehender Leistungsformate durch Dokumentationspflichten, Detektion oder strukturelle Prüfungsanpassungen. Diese pragmatischen Antworten adressieren zwar Integritätsfragen, lassen jedoch einen zentralen blinden Fleck sichtbar werden. Learning Analytics (LA) wird kaum systematisch genutzt, um KI-gestützte Lern- und Prüfungsprozesse empirisch zu verstehen und didaktisch weiterzuentwickeln. Vor dem Hintergrund des anhaltenden »Closing-the-loop«-Problems argumentiert der Beitrag, dass hybride Mensch-KI-Interaktionen neue Datenpotenziale eröffnen, etwa zu Prompting-Strategien, Iterationsschleifen oder ko-konstruktiver Problemlösung, die bislang unzureichend erschlossen sind. Der Beitrag schlägt hierfür das Rahmenkonzept AI-Integrated Learning Analytics (AILA) vor. Dieses Konzept integriert Learning Analytics, KI-gestützte Interventionen und pädagogische Zielsetzungen in geschlossenen Feedback-Schleifen. AILA versteht »Human Agency« und die Gestaltung von »Human–AI Interactions« als normativen Referenzpunkt und wird exemplarisch am Anwendungsfall »Assessment-as-Learning« illustriert.

Integrating GenKI and Learning Analytics: A Conceptual Framework for Interaction-Centered Learning Analytics

The growing adoption of generative artificial intelligence (GenAI) in schools and higher education is fundamentally changing the way assessment and feedback are conducted. While current debates mainly focus on safeguarding existing assessment formats, the potential of learning analytics (LA) to facilitate effective human–AI interactions remains largely unexplored. This paper addresses this gap by presenting an integrated, pedagogically grounded approach to learning analytics and generative AI. Assuming that GenAI is a new foundational technology that is reshaping learning environments, the paper introduces the AI-Integrated Learning Analytics (AILA) conceptual framework. AILA is conceived as an organizing framework

1 Basiert auf einer Keynote im Rahmen der Tagung.

for designing agency-oriented human–AI interactions, in which analytical data, AI-based interventions and pedagogical goals are linked within closed feedback loops. The framework is illustrated using the application scenario of assessment-as-learning. The paper serves as a conceptual starting point and outlines directions for future research and educational practice.

Ausgangspunkt

In den aktuellen Diskussionen um den Einsatz generativer KI (GenKI) in Schule und Hochschule dominieren zunächst Reaktionsmuster, die auf eine Absicherung bestehender Prüfungs- und Leistungsformate abzielen. Diese Debatten werden insbesondere durch nicht-beaufsichtigte, schriftliche Prüfungsformen herausgefordert, deren Validität und Zuschreibbarkeit von Eigenleistung durch den Einsatz von GenKI infrage gestellt werden (Holmes/Tuomi 2022; Kasneci et al. 2023). Entsprechend konzentrieren sich viele institutionelle Maßnahmen auf kompensatorische Strategien, etwa Transparenz- und Dokumentationspflichten zum KI-Einsatz (z.B. Prompt-Historien), den Einsatz technischer Detektionsverfahren, die Verlagerung hin zu mündlichen oder ergänzenden Prüfungsformaten unter Aufsicht (z.B. Kolloquien) sowie die Erweiterung schriftlicher Prüfungen um Portfolios und Reflexionsleistungen (Corbin et al. 2025; Debets et al. 2025). Empirische Studien und hochschulweite Policy-Analysen zeigen jedoch, dass insbesondere Verbote und Dokumentationsauflagen in offenen, nicht-beaufsichtigten Settings als praktisch kaum durchsetzbar eingeschätzt werden und daher zunehmend durch prüfungsdidaktische Anpassungen ersetzt werden, die weiterhin auf individuelle Leistungszurechnung abzielen (Gregorac et al. 2025).

Diese Vorschläge sind pragmatische Antworten auf die Herausforderung, die Integrität von Prüfungen und Leistungsnachweisen im Zeitalter generativer KI zu sichern. Zugleich offenbaren sie jedoch einen blinden Fleck: Learning Analytics (LA) wird dabei nicht mitgedacht. Obwohl LA prädestiniert wäre, die Nutzung von GenKI empirisch sichtbar zu machen, etwa durch die Analyse von Interaktionsdaten, die Erfassung von Lernverläufen oder die systematische Verknüpfung von Prozess- und Ergebnisdaten, bleibt dieser Zugang in der Debatte weitgehend unberücksichtigt.

Gerade in diesem Kontext tritt ein seit Langem bekanntes strukturelles Problem des Bildungssystems deutlicher hervor: das sogenannte Closing-the-loop-Problem (Clow 2012), also die fehlende systematische Rückkopplung zwischen Lernprozessen, Leistungsnachweisen und didaktischer Weiterentwicklung (Ifenthaler/Yau 2020; Wise 2014). Zwar entstehen in digitalen Lernumgebungen umfangreiche Prozess- und Interaktionsdaten, diese werden jedoch häufig weder zur formativen Unterstützung individueller Lernprozesse noch zur evidenzbasierten Weiterentwicklung von Lehr- und Prüfungsformaten genutzt (Dawson et al. 2019). Die zunehmende Nutzung generativer KI verschärft diese Problematik insofern, als

Lernprozesse verstärkt in hybriden Mensch-KI-Interaktionen stattfinden, deren Qualität, Tiefe und Entwicklung ohne analytische Verfahren weitgehend intransparent bleiben (Kasneci et al. 2023). Zugleich eröffnet GenKI neue, bislang kaum ausgeschöpfte Datenpotenziale, etwa zu Prompting-Strategien, Iterationsschleifen oder Formen ko-konstruktiver Problemlösung, die eine Schließung dieser Feedbackschleifen prinzipiell erleichtern könnten (Buckingham Shum et al. 2024). Wird auf eine systematische Integration von Learning Analytics verzichtet, bleibt damit weniger das strukturelle Problem selbst neu, sondern vielmehr eine zentrale didaktische Chance ungenutzt. GenKI-gestützte Lernprozesse lassen sich dann kaum über einzelne Anwendungsfälle hinaus systematisch verstehen, weiterentwickeln und lernwirksam gestalten. Während zahlreiche Studien die punktuelle Wirksamkeit generativer KI etwa als Tutoring-System oder zur Prüfungsvorbereitung belegen (Deng et al. 2025; Ansari et al. 2024), fehlt bislang ein integrierter analytischer Rahmen, der diese Effekte in Beziehung zu Lernprozessen, Kompetenzentwicklung und didaktischem Design setzt. An dieser Leerstelle – der Verbindung von GenKI-Nutzung, Learning Analytics und evidenzbasierter Gestaltung lernförderlicher Architekturen – setzt der vorliegende Beitrag an.

Ziel dieses Beitrags ist es, das Verhältnis von Learning Analytics (LA) und generativer Künstlicher Intelligenz (GenKI) im Bildungskontext neu zu bestimmen. Ausgangspunkt ist die Diagnose, dass in aktuellen Debatten zwar sowohl datenanalytische als auch technologiegetriebene Perspektiven präsent sind, diese jedoch meist parallel und nur selten in einer gemeinsam pädagogisch gerahmten Perspektive verhandelt werden. Während Learning Analytics primär auf die Analyse und Rückkopplung von Lernprozessen zielt, werden GenKI-Anwendungen häufig als eigenständige Werkzeuge zur Unterstützung von Lernen, Lehren oder Prüfen diskutiert. Vor diesem Hintergrund rückt der Beitrag eine offene Leitfrage in den Mittelpunkt: *Wie können Learning Analytics, generative KI und pädagogische Zielsetzungen so aufeinander bezogen werden, dass sie gemeinsam zur Gestaltung wirksamer Feedbackschleifen für Lehr- und Lernprozesse beitragen?* Die Frage zielt nicht auf eine abschließende Lösung, sondern auf eine konzeptionelle Rahmung, die das komplementäre Potenzial von LA und GenKI systematisch erschließt und zugleich pädagogische Zielsetzungen als normativen Bezugspunkt in den Mittelpunkt stellt.

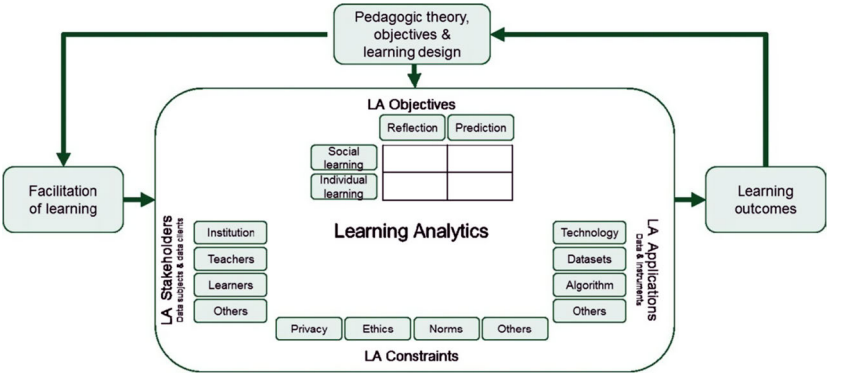
Der Beitrag ist wie folgt aufgebaut: Abschnitt 2 skizziert zentrale Grundannahmen und Grenzen etablierter Learning-Analytics-Frameworks als Ausgangspunkt. Abschnitt 3 führt das Rahmenkonzept AI-Integrated Learning Analytics (AILA) ein und leitet daraus pädagogische Designprinzipien ab. Abschnitt 4 illustriert die Logik des Frameworks am Anwendungsfall Assessment-as-Learning. Abschnitt 5 diskutiert offene Forschungs- und Gestaltungsfragen sowie Implikationen für die Bildungspraxis.

Theoretische Grundlagen: LA Framework als Ausgangspunkt

Learning Analytics (LA) wurde ursprünglich definiert als »the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs« (Siemens 2013: 1382). Früh einflussreiche Frameworks, wie das von Clow (2012), betonten die zyklische Struktur von Datenerhebung, Analyse, Feedback und Rückkopplung auf Lehr- und Lernprozesse (»Closing-the-loop Frameworks«). Zentral ist dabei die Grundidee, dass LA nicht als Selbstzweck betrieben wird, sondern als Feedback-System, das pädagogisch wirksam wird (Ferguson/Clow 2017).

In der Praxis zeigte sich jedoch, dass viele Implementierungen diesen Anspruch nur teilweise einlösen. Häufig konzentrieren sie sich auf die Bereitstellung von Visualisierungen oder Prognosen, etwa zur Vorhersage von Dropout-Risiken, ohne dass klar ersichtlich wird, wie diese Befunde tatsächlich in didaktische Handlungen übersetzt werden (Ifenthaler & Yau, 2020). Um diese Lücke zu schließen, haben Seufert et al. bereits 2019 ein theoriegeleitetes Modell aus pädagogischer Perspektive entwickelt, das Lernziele, Instruktionsdesign, Lehrhandlungen und Outcomes systematisch miteinander verknüpfen soll. Zugleich berücksichtigt es institutionelle und ethische Rahmenbedingungen und etabliert LA damit als ein pädagogisch fundiertes und normativ eingebettetes Konzept.

Abbildung 1: Design Framework for LA (Seufert et al. 2019)



Das »Design Framework for LA« (vgl. Abb. 1) hat vor diesem Hintergrund dazu beigetragen, LA als strukturierte Verbindung zwischen pädagogischen Zielen, Datenanwendungen und Stakeholdern zu etablieren. Besonders hervorzuheben ist die klare Differenzierung zwischen Zielen (z.B. Reflexion, Vorhersage), beteiligten Akteuren (Institutionen, Lehrpersonen, Lernende), technischen Anwendungen so-

wie normativen Rahmenbedingungen (z.B. Ethik, Datenschutz). Damit wird LA als geschlossenes System konzipiert, das Daten über Lernprozesse erhebt, aufbereitet und für pädagogische Entscheidungen nutzbar macht.

Mit dem Aufkommen generativer Künstlicher Intelligenz (GenKI) stößt dieses Modell jedoch zunehmend an konzeptionelle Grenzen. KI, insbesondere GenKI als Basistechnologie, erscheint darin allenfalls implizit als Teil der »Applications« und wird primär als technologische Komponente oder algorithmisches Werkzeug verstanden. Die qualitativ neuen Möglichkeiten von Large Language Models, etwa als Tutor, Coach oder Reflexionspartner in dialogischen Lernprozessen zu fungieren, sind im Rahmen des Modells bislang nicht explizit verankert (Holmes/Tuomi 2022; Sailer et al. 2024). Damit fehlt eine systematische Integration KI-basierter Interventionen in die pädagogische Logik des Frameworks. Zugleich wirkt die aktuelle Entwicklung generativer KI als Beschleuniger eines bereits zuvor angelegten Trends in der Learning-Analytics-Forschung: Durch Fortschritte in Machine Learning, Natural Language Processing und multimodalen Datenquellen lassen sich Lernprozesse nicht nur feingranularer erfassen, sondern auch in ihrer zeitlichen, dialogischen und kontextuellen Dynamik interpretieren (Sailer et al. 2024). Gerade die Interaktionsdaten aus GenKI-gestützten Lernsettings verschärfen damit die Frage nach dem Theory-Evidence Alignment, da bestehende theoretische Modelle häufig nicht hinreichend ausdifferenziert sind, um diese neuen Formen ko-konstruktiver Mensch-KI-Interaktion konzeptionell zu fassen. Entsprechend gewinnt ein iteratives Verständnis an Bedeutung, in dem Theorien nicht nur das Design datenbasierter Systeme leiten, sondern durch empirische Evidenz aus GenKI-Interaktionen kontinuierlich überprüft und weiterentwickelt werden (Wong et al. 2019).

Diese Entwicklung verdeutlicht, dass ein rein analytics-zentriertes Modell nicht mehr ausreicht, um die Potenziale und Herausforderungen generativer KI im Bildungskontext adäquat zu erfassen. Klassische LA-Frameworks fokussieren vor allem die Gewinnung und Aufbereitung von Daten; KI-basierte Systeme hingegen erfordern eine doppelte Perspektive: Sie sind nicht nur Analyseinstrumente, sondern zugleich Akteure im Lernprozess, die aktiv in Interaktionen eingreifen, Feedback geben oder Lernhandlungen initiieren. Daraus ergeben sich zwei zentrale Konsequenzen für eine Weiterentwicklung:

1. Integration statt Separation: Anstelle getrennter Diskurse zu Learning Analytics und KI braucht es ein Rahmenmodell, das beide Stränge systematisch verbindet.
2. Pädagogische Orientierung: Die Integration muss lernzentriert erfolgen. Ausgangspunkt sind pädagogische Zielsetzungen, in die sowohl Analytics als auch KI-gestützte Interventionen eingebettet werden.

Vor diesem Hintergrund schlagen wir das Konzept der AI-Integrated Learning Analytics (AILA) vor. Es erweitert das klassische Closed-Loop-Modell um eine Ebene KI-gestützter Interventionen, die unmittelbar mit Lernzielen, pädagogischen Orientierungen und Human Agency verknüpft werden. Human Agency im Rahmen von AILA bezeichnet sowohl eine individuelle Fähigkeit von Lernenden und Lehrenden als auch ein zentrales Gestaltungsprinzip KI-gestützter Lern- und Analysesysteme. Auf individueller Ebene beschreibt Human Agency die Fähigkeit, KI-basierte Analysen und Interventionen nicht passiv zu übernehmen, sondern diese aktiv zu reflektieren, zu bewerten und in Übereinstimmung mit eigenen Zielen, Kontexten und Verantwortlichkeiten zu gestalten. Auf systemischer Ebene zielt Human Agency darauf ab, Lernumgebungen so zu gestalten, dass diese Handlungsspielräume nicht unterlaufen, sondern gezielt erhalten und erweitern. Dazu gehört insbesondere, dass Lernende und Lehrende die Entscheidungsfreiheit behalten, in welchen Bereichen sie Aufgaben an KI delegieren, in welchen sie KI zur Erweiterung eigener Fähigkeiten nutzen und in welcher Weise sie KI-basierte Empfehlungen aufnehmen, anpassen oder bewusst zurückweisen. Pädagogische Verantwortung bleibt damit nicht an das KI-System delegiert, sondern ist strukturell und normativ an die Human Agency der beteiligten Akteure gekoppelt (vgl. hierzu auch Human Agency als Leitprinzip des AI-Literacy-Rahmenkonzepts der OECD 2025).

Konzeptionelle Idee: AI-Integrated Learning Analytics (AILA)

Überblick über das Rahmenkonzept AILA

Vor diesem Hintergrund wird im Beitrag das Konzept der AI-Integrated Learning Analytics (AILA) skizziert. AILA zielt darauf ab, analytische Daten, KI-gestützte Interventionen und pädagogische Zielsetzungen in einer geschlossenen Feedback-Schleife zusammenzuführen. Im Unterschied zu klassischen Learning-Analytics-Ansätzen, die häufig bei der Bereitstellung von Daten oder Visualisierungen stehen bleiben, versteht AILA generative KI als aktiven Bestandteil des Lernprozesses.

Zentral ist dabei ein relationales Verständnis von Human Agency. Lernende und Lehrende sollen KI-basierte Rückmeldungen nicht passiv übernehmen, sondern reflektieren, bewerten und in eigene Lern- und Lehrentscheidungen integrieren können. Dieses Verständnis knüpft an aktuelle Arbeiten an, die Human Agency im Kontext generativer KI nicht als individuelles Merkmal allein, sondern als Ergebnis der Gestaltung soziotechnischer Systeme verstehen (Krakowski 2025; OECD 2025).

Vor diesem Hintergrund versteht sich AILA nicht primär als daten- oder technologiezentriertes Framework, sondern als pädagogisch fundierter Ordnungsrahmen, dessen Kern in der Gestaltung von Human-AI Interactions unter Wahrung

und Stärkung von Human Agency liegt. Learning Analytics und generative KI werden dabei nicht als getrennte Ebenen konzipiert, sondern als miteinander verschränkte Komponenten, die Interaktionen strukturieren, Lernhandlungen beeinflussen und Reflexionsprozesse ermöglichen. AILA rückt somit nicht Daten oder Algorithmen, sondern die qualitative Gestaltung von Mensch-KI-Interaktionen in den Mittelpunkt und versteht Human Agency als normativen Referenzpunkt, an dem Analyse, Intervention und Systemgestaltung auszurichten sind (Kasneci et al. 2023; Buckingham Shum et al. 2024).

Darüber hinaus ist GenKI in diesem Kontext nicht lediglich als weiteres pädagogisches Werkzeug zu verstehen, sondern als neue Basistechnologie digitaler Lernumgebungen. Ähnlich wie Learning-Management-Systeme oder Learning-Analytics-Infrastrukturen wirkt GenKI quer zu einzelnen Anwendungsfällen, indem sie zentrale Funktionen der Wissensrepräsentation, der Interaktion und der Adaptivität neu organisiert. Der vorgeschlagene Integrationsansatz reagiert somit auf eine strukturelle Veränderung der Lernumgebung, in der analytische, pädagogische und normative Perspektiven von Beginn an gemeinsam gedacht werden müssen (Holmes/Tuomi 2022).

Abbildung 2: AI-Integrated Learning Analytics (AILA) – konzeptioneller Ordnungsrahmen

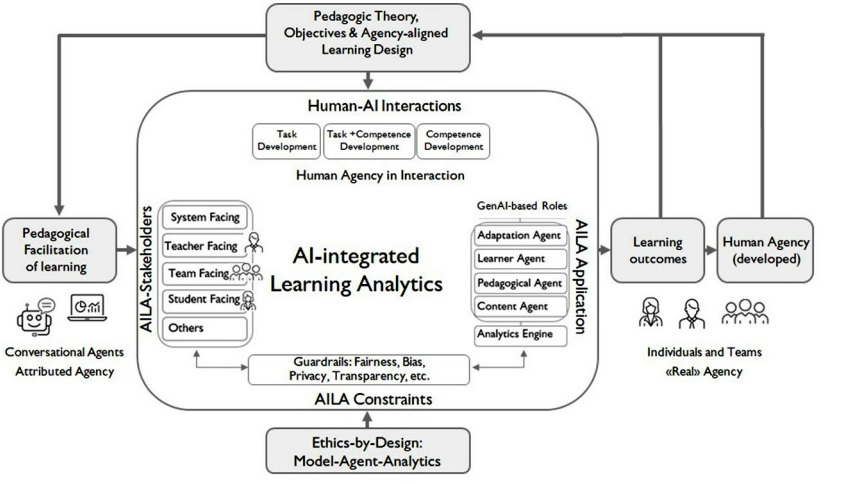


Abbildung 2 skizziert das AILA-Framework als konzeptionellen Ordnungsrahmen zur Integration von Learning Analytics und generativer KI. Die Darstellung versteht sich nicht als abgeschlossenes Modell, sondern als heuristische Struktur, die zentrale Gestaltungsdimensionen, insbesondere Human-AI Interactions, Human Agency sowie ethische und pädagogische Rahmungen systematisch in Beziehung

setzt. Sie dient damit als Ausgangspunkt für weitere theoretische Präzisierungen und empirische Untersuchungen in unterschiedlichen Anwendungskontexten.

Anwendungsbeispiel: Assessment-as-Learning im AILA-Framework

Aufbauend auf der in Abschnitt 2 eingeführten Logik von Assessment-as-Learning zeigt das folgende Anwendungsbeispiel, wie diese Perspektive im AILA-Framework konkret umgesetzt werden kann. Assessment wird dabei nicht als isoliertes Ereignis verstanden, sondern als wiederkehrender, lernsteuernder Prozess, in dem diagnostische Informationen, Feedback und Lernunterstützung kontinuierlich miteinander verschränkt sind. Diese Perspektive ist lernpsychologisch gut fundiert und wird insbesondere durch Arbeiten zum Test-Enhanced Learning gestützt, die zeigen, dass wiederholte Abruf- und Testprozesse nicht nur der Leistungsüberprüfung dienen, sondern selbst einen substanziellen Beitrag zur langfristigen Wissensstabilisierung leisten (Roediger/Karpicke 2006).

Abbildung 3 veranschaulicht, wie dieses Verständnis von Assessment-as-Learning innerhalb des AILA-Frameworks umgesetzt werden kann. Das dargestellte Szenario zeigt kein einzelnes KI-Werkzeug, sondern ein koordiniertes Zusammenspiel pädagogischer Rollen, die durch ein KI-gestütztes Multi-Agenten-System realisiert werden. Die Agenten sind dabei nicht als autonome Akteure zu verstehen, sondern als funktionale Ausprägungen eines integrierten Lern- und Analysesystems.

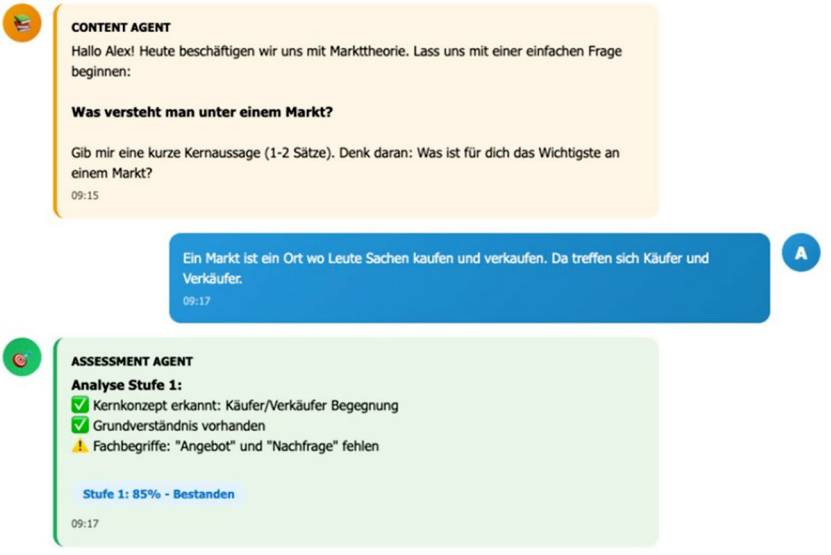
Der Pedagogical Agent übernimmt in diesem Szenario temporär die Rolle eines Assessment Agent, indem der pädagogische Fokus gezielt auf diagnostische, rückmeldungsbezogene und kompetenzorientierte Funktionen ausgerichtet wird. Er strukturiert Aufgabenformate, definiert Bewertungskriterien und steuert formative sowie summative Analyseprozesse. Der Content Agent kuratiert und stellt Aufgaben adaptiv bereit, wobei Schwierigkeitsgrad, Aufgabenvariation und inhaltliche Schwerpunkte an den jeweiligen Lernstand angepasst werden.

Die Lernantworten werden vom assessment-fokussierten Pedagogical Agent analysiert, Fortschritte entlang von Kompetenzmodellen gespiegelt und typische Fehlkonzepte identifiziert. Parallel dazu aktualisiert der Learner Agent kontinuierlich Lernprofile, dokumentiert Entwicklungsverläufe und leitet Indikatoren zur Selbstregulation (*self-regulated learning*, SRL) ab. Diese analytischen Informationen bilden die Grundlage für den Adaptation Agent, der Scaffolds, Feedbackformen und Lernpfade situativ anpasst.

Das Zusammenspiel der Agenten – wie in Abbildung 3 dargestellt – realisiert einen geschlossenen Feedback-Loop, in dem Analyse, Intervention und Lernhandlung eng miteinander verschränkt sind. Zentrale Funktion dieses Loops ist nicht die

Automatisierung von Bewertung, sondern die lernwirksame Rückführung diagnostischer Informationen in den Lernprozess.

Abbildung 3: Darstellung eines Beispiels für ein Multi-Agenten System



Im Sinne von AILA bleibt dabei Human Agency der normative Referenzpunkt. Lernende erhalten nicht nur automatisierte Rückmeldungen, sondern dialogisch aufbereitete Informationen, die zur Reflexion eigener Lernstrategien, zur Einschätzung des Kompetenzstands und zur bewussten Steuerung weiterer Lernschritte anregen sollen. Lehrende behalten die Möglichkeit, Analyseergebnisse zu interpretieren, pädagogische Entscheidungen zu treffen und das System gezielt zu justieren.

Assessment wird in diesem Szenario somit nicht als nachgelagerter Kontrollmechanismus verstanden, sondern als integraler Bestandteil eines lernförderlichen Designs, das Kompetenzentwicklung aktiv anleitet und unterstützt. Das Beispiel verdeutlicht exemplarisch, wie AILA die Potenziale von Learning Analytics und generativer KI bündelt, um Assessment-as-Learning als kontinuierlichen, agentivitätsorientierten Prozess zu realisieren.

Diskussion und Ausblick

Der Beitrag versteht sich als konzeptioneller Impuls, der zeigt, wie Learning Analytics und generative KI im Sinne eines *Assessment-as-Learning* systematisch zusam-

mengedacht werden können. Der Mehrwert liegt weniger in einzelnen Technologien als in der pädagogisch begründeten Integration von Analyse, KI-gestützter Intervention und Reflexion innerhalb geschlossener Feedback-Schleifen. Mit dem AI-LA-Framework und den daraus abgeleiteten Designprinzipien wird eine Perspektive vorgeschlagen, in der *Human Agency* und die bewusste Gestaltung von *Human–AI Interactions* den normativen Referenzpunkt bilden.

Die entwickelten Designprinzipien markieren zugleich zentrale Verschiebungen gegenüber etablierten Learning-Analytics-Ansätzen. Insbesondere die interaktionszentrierte Ausrichtung von Analytik, die konsequente Verankerung von Closed-Loop-Architekturen in pädagogischen Zielsetzungen sowie die Betonung reflexiver statt direktiver Nutzung analytischer Rückmeldungen adressieren bekannte Limitationen analytics-zentrierter Modelle. AILA positioniert Learning Analytics damit nicht als nachgelagerte Evidenzinstanz, sondern als integralen Bestandteil lernwirksamer Interaktionen, in denen generative KI aktiv an der Strukturierung von Lernprozessen beteiligt ist, ohne menschliche Urteilshoheit zu ersetzen.

Zugleich bleiben zentrale Fragen offen. Es bedarf empirischer Forschung, um zu untersuchen, wie sich agency-orientierte, AI-integrierte Learning-Analytics-Architekturen auf Lernprozesse, Kompetenzentwicklung und Selbstregulation auswirken. Offene Forschungsfragen betreffen unter anderem die Validität interaktionsbasierter Indikatoren, die Gestaltung transparenter und revidierbarer KI-Interventionen sowie mögliche Spannungsfelder zwischen Unterstützung und Übersteuerung. Zu klären ist zudem, unter welchen Bedingungen Learning Analytics tatsächlich reflexive Sinnbildungsprozesse fördern, anstatt neue Formen impliziter Steuerung zu etablieren.

Darüber hinaus weisen die Designprinzipien über den exemplarischen Anwendungsfall des *Assessment-as-Learning* hinaus. AILA ist grundsätzlich auf unterschiedliche didaktische Kontexte übertragbar, etwa auf interaktionsorientierte Kompetenzentwicklung in der Lehrpersonenbildung, auf metakognitiv ausgerichtete Aufgabenformate oder auf kritisch-reflexive Unterstützungsformen im Umgang mit KI-generierten Inhalten. Entscheidend ist dabei nicht das konkrete Anwendungsszenario, sondern die Art der Gestaltung von Learning Analytics und KI-gestützten Interventionen: Sie sollten Interaktionsprozesse sichtbar machen, Reflexion anregen und die menschliche Entscheidungshoheit gezielt unterstützen.

Schließlich ergeben sich Implikationen für Governance und Systemgestaltung. Wenn Learning Analytics und generative KI als strukturbildende Komponenten von Lernumgebungen verstanden werden, müssen ethische und governancebezogene Anforderungen als Designbedingungen in Feedback-Architekturen integriert werden. Transparenz, Datenschutz, Fairness und Verantwortlichkeit sind dabei nicht nur normative Leitlinien, sondern beeinflussen konkret, welche Daten erhoben werden, wie analytische Modelle interpretiert werden dürfen und in welcher Weise

KI-gestützte Interventionen hinterfragt, angepasst oder zurückgewiesen werden können.

Der Beitrag plädiert damit für eine pädagogisch fundierte Neujustierung, in der Learning Analytics und generative KI als sich wechselseitig ermöglichende Bausteine einer lernförderlichen Assessment- und Feedback-Architektur verstanden werden. AILA bietet hierfür einen konzeptionellen Rahmen, der Design, Analyse und pädagogische Verantwortung systematisch verbindet und zugleich Anschlussstellen für weiterführende empirische Forschung und gestaltungsorientierte Entwicklung eröffnet.

Literatur

Ansari, Aisha N./Ahmad, Sohail/Bhutta, Sadia M. (2024): »Mapping the global evidence around the use of ChatGPT in higher education: A systematic scoping review«, in: Education and Information Technologies, 29(9), S. 11281–11321.

Buckingham Shum, Simon/Ferguson, Rebecca/Martinez-Maldonado, Roberto (2024): »Human-centred learning analytics«, in: Journal of Learning Analytics, 11(1), S. 1–15.

Clow, Dough (2012): »The learning analytics cycle: Closing the loop effectively«, in: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM (LAK ›12), S. 134–138.

Corbin, T./Dawson, P./Liu, D. (2025): »Talk is cheap: Why structural assessment changes are needed for a time of GenAI«, in: Assessment & Evaluation in Higher Education, 50(7).

Cukurova, M. (2025): »Human–AI interaction and learning: Rethinking educational design for generative AI«, in: Computers & Education, 203, 104870.

Dawson, Shane/Joksimović, Srecko/Poquet, Oleksandra/Siemens, George (2019): »Increasing the impact of learning analytics«, in: Proceedings of the Ninth International Conference on Learning Analytics & Knowledge, ACM (LAK ›19)

Debets, Tim/Banihashem, Seyyed K./Joosten-Ten Brinke, Desirée/Vos, Tanja E./Maillette de Buy Wenniger, Gideon/Camp, Gino (2025): »Chatbots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts«, in: Computers & Education, 234, 105323.

Deng, Ruiqi/Jiang, Maoli/Yu, Xinlu/Lu, Yuyan/Liu, Sasha (2025): »Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies«, in: Computers & Education, 227, 105224.

Dignum, Virginia (2019): Responsible artificial intelligence. How to develop and use AI in a responsible way. Springer.

Doshi, T./Hauser, D. (2024): »Overreliance on AI advice: Cognitive and behavioral consequences«, in: Nature Human Behaviour, 8, S. 112–123.

Ferguson, R./Clow, D. (2017): »Where is the evidence? A call for learning analytics evidence«, in: Journal of Learning Analytics, 4(2), S. 4–8.

Floridi, Luciano/Cowls, Josh/Beltrametti, Monica/Chatila, Raja/Chazerand, Patrice/Dignum, Virginia/Luetge, Christoph/Madelin, Robert/Pagallo, Ugo/Rossi, Francesca/Schafer, Burkhard/Valke, Peggy/Vayena, Effy (2018): »AI4People – An ethical framework for a good AI society: Opportunities, Risks, Principles, and Recommendations«. in: Minds and Machines, 28, S. 689–707.

Gašević, Dragan/Kovanović, Vitomir/Joksimović, Srećko (2017): »Piecing the learning analytics puzzle: A consolidated model of a field of research and practice«. Learning: Research and Practice, 3(1), S. 63–78.

Gregorac, Ana/Brünner, Benedikt/Ebner, Martin (2025): »Chatbots in education: A systematic rapid literature review«, in: R. H. J. Cohen (Ed.), Proceedings of the SITE Conference, S. 652–657.

Greller, Wolfgang/Drachsler, Hendrik (2012): »Translating learning into numbers: A generic framework for learning analytics«, in: Journal of Educational Technology & Society, 15(3), S. 42–57.

Guo, Y./Li, J./Sun, Y. (2025): »Human–AI collaboration in learning: Agency, trust, and performance«, in: Learning and Instruction, 85, 101785. https://doi.org/10.1016/j.learninstruc.2024.101785.

Holmes, W./Porayska-Pomsta, K. (2023): »Ethics, equity, and the design of AI in education«, in: British Journal of Educational Technology, 54(6), S. 2131–2147. https://doi.org/10.1111/bjet.13368.

Holmes, W./Tuomi, I. (2022): »State of the art in AI and education« (OECD Education Working Papers No. 270), in: OECD Publishing. https://doi.org/10.1787/ca39a3e5-en

Ifenthaler, Dirk/Yau, Jane Y.-K. Y. (2020): »Utilising learning analytics to support study success in higher education: A systematic review«, in: Educational Technology Research and Development, 68(4), S. 1961–1990.

Jobin, Anna/Ienca, Marcello/Vayena, Effy (2019): »The global landscape of AI ethics guidelines«, in: Nature Machine Intelligence, 1, S. 389–399.

Kasneci, Eenkelejda/Sessler, Kathrin/Küchemann, Stefan/Bannert, Maria/Dementieva, Daryna/Fischer, Frank/Gasser, Urs/Groh, Georg/Günnemann, Stephan/Hüllermeier, Eyke/Krusche, Stephan/Kutyniok, Gitta/Michaeli, Tilman/Nerdel, Claudia/Pfeffer, Jürgen/Poquet, Oleksandra/Sailer, Michael/Schmidt, Albrecht/Seidel, Tina/Stadler, Matthias/Weller, Jochen/Kuhn, Jochen/Kasneci, Gjergji (2023): »ChatGPT for good? On opportunities and challenges of large language models for education«, in: Learning and Individual Differences, 103, 102274.

Krakowski, Sebastian (2025): »Human–AI agency in the age of generative AI«, in: Information and Organization, 35(1), 100560.

Lockyer, Lori/Heathcote, Elizabeth/Dawson, Shane (2013): »Informing pedagogical action: Aligning learning analytics with learning design«, in: American Behavioral Scientist, 57(10), S. 1439–1459.

OECD (2025): Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft). OECD Publishing. https://ailiteracyframework.org

Pardo, Abelardo/Siemens, George (2014): »Ethical and privacy principles for learning analytics«, in: British Journal of Educational Technology, 45(3), S. 438–450.

Roediger, Henry L./Karpicke, Jeffrey D. (2006): »Test-enhanced learning: Taking memory tests improves long-term retention«, in: Psychological Science, 17(3), S. 249–255.

Sailer, Michael/Ninaus, Manuel/Huber, Stefan E./Bauer, Elisabeth/Greiff, Samuel (2024): »The end is the beginning is the end: The closed-loop learning analytics framework«, in: Computers in Human Behavior, 158, 108305.

Seufert, Sabine/Meier, Christoph/Soellner, Matthias/Rietsche, Roman (2019): »A pedagogical perspective on big data and learning analytics: A conceptual model for digital learning support«, in: Technology, Knowledge and Learning, 24(4), S. 599–619.

Siemens, George (2013): »Learning analytics: The emergence of discipline«. in: American Behavioral Scientist, 57(10), S. 1380–1400.

Slade, Sharon/Prinsloo, Paul (2013): »Learning analytics: Ethical issues and dilemmas«. in: American Behavioral Scientist, 57(10), S. 1510–1529.

Wise, Alyssa F. (2014): »Designing pedagogical interventions to support student use of learning analytics«, in: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, ACM (LAK ›14), S. 203–211.

Wong, Jaqueline/Baars, Martine/de Koning, Björn B./van der Zee, Tim/Davis, Dan/Khalil, Mohammad/Houban, Geert-Jan/Paas, Fred (2019): »Educational theories and learning analytics: From data to knowledge«, in: Dirk Ifenthaler/Dana-Kristin Mah/Y.K. Y. Jane (Hg.), Utilizing learning analytics to support study success, S. 3–25. Springer.

Künstliche Intelligenz

Von Betroffenen zu Beteiligten

Ein studentisch entwickeltes Dialogformat zu KI in der Hochschullehre[1]

Sarah Becker[2] und Katharina Westphal[3]

Künstliche Intelligenz verändert Hochschulen tiefgreifend. Studierende sind davon unmittelbar betroffen, bleiben in hochschulstrategischen Diskursen jedoch oft unsichtbar. Dieser Beitrag stellt ein Workshopformat vor, das im Rahmen der Konferenz Learning AID 2025 erprobt wurde, und Studierende, Lehrende sowie strategische Entscheider:innen in einen strukturierten Dialog zum Umgang mit KI an Hochschulen brachte. Methodisch kombiniert das Format die kollegiale Fallberatung und das World Café und verortet sich damit zwischen Dialog und Mitgestaltung (vgl. Ditzel 2025). Es zeigt, wie Partizipation nicht nur gefordert, sondern praktisch gestaltet werden kann. Die Diskussionen machten sichtbar: Eine Misstrauenskultur prägt die Haltung gegenüber Studierenden, Regelungen bleiben uneinheitlich, Zugänge zu KI sind ungleich verteilt und Qualifizierungsangebote lückenhaft. Zugleich entstanden konkrete Lösungsideen. Damit wird deutlich: Partizipative Formate sind notwendig, um KI in der Hochschullehre gerecht und zukunftsfähig zu gestalten.

From affected parties to participants: A dialogue format on AI in higher education teaching, developed by students.

Artificial intelligence is profoundly changing higher education. Students are directly affected by this, yet they often remain invisible in strategic higher education discourse. This article presents a workshop format that was tested at the Learning AID 2025 conference and brought students, teachers and decision-makers together in a structured dialogue. Methodologically, the format combines collegial case consultation and World Café, thus positioning itself between dialogue and co-creation (cf. Ditzel 2025). It shows that participation is not only demanded but can also be shaped in practical terms. The discussions revealed that a culture of mistrust shapes attitudes towards students, regulations remain inconsistent, access to AI is unevenly distributed, and training opportunities are fragmentary. At the same time, concrete ideas for

1 Basiert auf einem Workshop im Rahmen der Tagung.

2 ORCID: 0009–0005-6390-1556

3 ORCID: 0009–0003-3653-5269

solutions emerged. This makes it clear that participatory formats are necessary to ensure that AI in higher education is both fair and sustainable.

Einleitung

Die Entwicklungen im Bereich Künstlicher Intelligenz (KI) prägen die Hochschulen in vielfältiger Weise. Diskussionen drehen sich um didaktische Konzepte, Prüfungsformate oder ethische Fragen, von der Gestaltung KI-sicherer Prüfungen bis hin zu Fragen zu Bildungsgerechtigkeit und digitaler Souveränität. Diese Entwicklungen haben das Potenzial, den Studienalltag und das System Hochschule grundlegend zu verändern. Studierende sind von diesen Veränderungen unmittelbar betroffen. Ihre Stimmen bleiben jedoch in den meisten Strategie- und Entscheidungsprozessen ungehört oder werden, wenn überhaupt, nur punktuell eingebunden. Der KI-Monitor 2025 des Hochschulforum Digitalisierung (HFD) verdeutlicht dies: An keiner der befragten Universitäten werden Studierende stark in solche Prozesse eingebunden, an HAWs nur zu zehn Prozent (Budde/Tobor 2025: 14).

Die Frage nach der Partizipation von Studierenden rückt damit ins Zentrum. Sie gilt längst als Indikator für Qualität und Zukunftsfähigkeit von Hochschulen (Dietrich/Ditzel/Rasch/Schrader 2025: 1), bleibt in der Praxis aber häufig auf symbolische Formate wie Evaluationen oder Gremiensitze beschränkt, was als »Schein-Partizipation« kritisiert wird (Biehl 2024: 140). Unser Workshop setzt genau an dieser Stelle an: Er durchbricht diese Logik, indem er auf statusgruppenübergreifenden Dialog und konkrete Ansätze zur Mitgestaltung setzt. Ziel ist es, studentische Stimmen nicht nur hörbar zu machen, sondern wirksam in hochschulische Aushandlungsprozesse einzubringen: ein Dialogformat von und mit Studierenden.

Hintergrund: Der KI-Sprint der DigitalChangeMaker im Hochschulforum Digitalisierung

Der Workshop entstand im Rahmen des KI-Sprints der DigitalChangeMaker-Initiative des Hochschulforum Digitalisierung (HFD), einer vom HFD 2018 ins Leben gerufenen Initiative zur Stärkung der studentischen Partizipation an der digitalen Transformation von Hochschulen. Seit Ende 2023 entwickeln Studierende im Rahmen des KI-Sprints Positionen zur Rolle von KI in der Hochschulbildung und bringen diese in den hochschulischen Diskurs ein. 2024 wurden vier zentrale Dimensionen formuliert, die 2025 durch Blogbeiträge weiter geschärft wurden, und als Grundlage für den Workshops dienten.

Eine erste Dimension betonte die Bedeutung von *Haltung und Partizipation*: Statt einer Misstrauenskultur braucht es Vertrauensräume, in denen studentische Per-

spektiven systematisch eingebunden werden (Becker & Biehl 2025a). Eine zweite Dimension widmete sich den *Regelungen* und forderte klare, transparente Leitlinien anstelle uneinheitlicher Vorgaben, die Unsicherheit bei Studierenden wie Lehrenden erzeugen (Becker & Biehl 2025b). Drittens rückte das Thema *Teilhabe und Zugänge* in den Blick: Ungleiche Voraussetzungen beim Zugang zu Tools und Kompetenzen sollen abgebaut werden, zugleich gilt es, das Recht auf Nichtnutzung abzusichern (Gostmann 2025). Schließlich betonte eine vierte Dimension die Notwendigkeit von *Qualifizierung*, insbesondere die curriculare Verankerung von AI Literacy und Future Skills sowie den Ausbau von Peer-Formaten, in denen studentische Multiplikator:innen wie KI-Tutor:innen eine zentrale Rolle spielen (Becker/Biehl 2025c).

Diese Dimensionen rahmten die Fallbeispiele des Workshops und machten die abstrakten Positionen für die Teilnehmenden greifbar.

Workshopdesign: Von der Idee zur Umsetzung

Theoretischer Bezug

Für Ditzel (2025: 20) bedeutet Partizipation nicht nur Teilhabe, sondern die aktive Gestaltung von Rahmenbedingungen in der Hochschulbildung. Er unterscheidet dabei vier Stufen: *Informieren*, *Feedback*, *Dialog* und *Mitgestalten*. Während die ersten beiden eher unidirektionale Kommunikation beschreiben, eröffnet *Dialog* wechselseitigen Austausch, ohne notwendigerweise die Praxis zu verändern. Erst *Mitgestaltung* ermöglicht eine echte Veränderung von Praxis, sodass Studierende aktiv an Entscheidungen und der Weiterentwicklung von Praktiken beteiligt sind. Das Workshopformat bewegt sich in dieser Logik zwischen *Dialog* und *Mitgestaltung*: Es schafft Räume für Verständigung und zielt zugleich auf *Co-Creation* ab (Bovill 2020), indem statusgruppenübergreifend konkrete Lösungen entwickelt werden, die über den KI-Sprint und das HFD in hochschulpolitische Diskurse zur Gestaltung von KI in der Hochschulbildung einfließen.

Methodische Einordnung und Ablauf

Das Dialogformat kombinierte zwei methodische Ansätze: die *Kollegiale Fallberatung* und das *World Café*, um Studierende, Lehrende und Entscheider:innen in einen gleichrangigen Austausch zu bringen und gemeinsam Lösungsansätze zu entwickeln. Die Kollegiale Fallberatung stellt ein etabliertes Verfahren dar, das auf Gleichrangigkeit, Rollenwechsel und gegenseitiger Beratung basiert, sodass die Teilnehmenden als »Peers auf Zeit« zusammenarbeiteten (vgl. Hochschulforum Digitalisierung 2025a; Bosse/Lübcke 2024). Das World Café ergänzt diesen Ansatz,

indem es die systematische Durchmischung der Perspektiven ermöglicht und die inhaltlichen Zusammenhänge zwischen den Dimensionen hervorhebt.

Der fünfstufige Ablauf wurde bewusst so gestaltet, dass er schrittweise von Orientierung über gemeinsame Analyse hin zu Co-Kreation und Transfer führt:

Phase:

1. Einführung in die vier Dimensionen des KI-Sprints anhand von Fallbeispielen
2. Kleingruppenarbeit mit kollegialer Fallberatung: Analyse der Fälle und erste Lösungsansätze
3. Kurzpräsentationen im Plenum
4. Perspektivenmischung im World Café: neue Gruppenkonstellationen verknüpften Ideen und entwickelten diese weiter.
5. Blitzlicht: Transferfrage »Was nehme ich an meine Hochschule mit?«

Damit verbindet das Format zwei Logiken: Es sichert Gleichrangigkeit im Austausch durch die Kollegiale Fallberatung und fördert durch das World Café die systematische Durchmischung. Zugleich folgt es dem Anspruch, Partizipation nicht bei Information oder Feedback enden zu lassen, sondern bewusst in Richtung Dialog und Mitgestaltung zu führen.

Nach der ersten Kleingruppenarbeit werden die Teilnehmenden neu zusammengesetzt, wodurch Personen aus unterschiedlichen Gruppen aufeinandertreffen, ihre Ergebnisse austauschen und diese gemeinsam weiterentwickeln. Dafür wurden im Vorfeld Fallbeispiele aus verschiedenen Dimensionen miteinander kombiniert, die inhaltlich eng verbunden sind. Ein Beispiel dafür ist die Verknüpfung von Haltung und Regelungen: Ohne eine vertrauensvolle Basis entstehen Regeln oft auf Basis von Misstrauen, was Studierende verunsichert, und eine offene Lernkultur verhindern kann. Umgekehrt können transparente, partizipativ entwickelte Regeln Vertrauen erst ermöglichen. Haltung und Regelungen sind also wechselseitig aufeinander angewiesen (vgl. Becker/Biehl 2025a, 2025b; Ditzel 2025).

Vorläufige Ergebnisse und Ableitungen des Workshops

Der Workshop zeigte, dass im strukturierten Dialog zwischen Studierenden, Lehrenden und Entscheider:innen Unterschiede in Erfahrungen, Erwartungen und Rahmenbedingungen sichtbar, und zugleich bearbeitbar werden. Die Diskussionen verdichteten sich entlang der vier Dimensionen des KI-Sprints, ergänzt durch die querliegende Perspektive einer Beziehungskultur als Gelingensbedingung.

Haltung und Partizipation. Lehrende und Entscheider:innen berichteten, studentische Sorgen bislang kaum zu kennen. Studierende schilderten hingegen, beim Einsatz von KI häufig unter Generalverdacht zu stehen. Im Austausch entstand die Idee, Fragen des Umgangs mit KI, insbesondere Vertrauen, Transparenz und Fehlerkultur, nicht nur fallbezogen, sondern hochschulweit und statusgruppenübergreifend zu thematisieren und Gespräche zwischen den Statusgruppen durch kompakte Leitfäden zu unterstützen.

Regelungen. Teilnehmende beschrieben inkonsistente Regeln innerhalb von und über Studiengänge hinweg: KI sei in manchen Veranstaltungen erlaubt, in anderen strikt verboten, teils sogar innerhalb eines Moduls. Diese Uneinheitlichkeit führe zu Unsicherheit und Misstrauen (Becker & Biehl 2025b). Als Lösungsansatz wurde vorgeschlagen, Regeln partizipativ zu entwickeln und hochschulweit abzustimmen, sie transparent zu veröffentlichen und regelmäßig zu überprüfen, etwa im Rahmen von Pilotphasen und Evaluationszyklen.

Teilhabe und Zugänge. Ungleiche Vorerfahrungen und Ressourcen erschweren den Einstieg in KI, zugleich kann KI Barrieren abbauen (Hadar Shoval 2025). Zentrale Forderungen aus dem Workshop waren das Recht auf Nichtnutzung sowie transparente Alternativen bei Prüfungs- und Studienleistungen.

Qualifizierung. Studierende berichteten, dass sie ihre KI-Kompetenzen autodidaktisch erwerben. Das deckt sich mit aktuellen Befragungsergebnissen: Mehr als die Hälfte der Befragten im CHE-CHECK bewertet das Qualifizierungsangebot ihrer Hochschule als unzureichend (Hüsch/Horstmann/Breiter 2024). Einigkeit bestand darin, Qualifizierung curricular zu verankern, Zertifikatsangebote zu schaffen und Peer-Formate mit studentischen Multiplikator:innen auszubauen.

Beziehungsarbeit als verbindende Dimension

Über alle Dimensionen hinweg zeigte sich, dass Vertrauen, Fehlerkultur und transparente Kommunikation zentrale Voraussetzungen sind. Lehrende müssen nicht »allwissend« auftreten, Studierende können Expertise einbringen, ohne automatisch in Verdacht zu geraten, etwa unerlaubte KI-Nutzung oder Regelverstöße zu begehen. Erst diese Haltung verschiebt Partizipation von symbolischer Einbindung hin zu Dialog und Mitgestaltung.

Spannungsfelder und Wege der Auflösung

Im Verlauf der Fallarbeit und der statusgruppenübergreifenden Diskussionen traten kristallisierten sich wiederkehrende Spannungsfelder heraus. Besonders sicht-

bar wurde das Spannungsfeld zwischen Kontrolle und Vertrauen. Statt allein auf »KI-sichere Prüfungen« zu setzen, plädierten Teilnehmende für transparente Verfahren und faire Maßstäbe. Weitere Spannungsfelder lagen zwischen Einheitlichkeit und Fachspezifik, Innovationstempo und Rechtssicherheit sowie Verpflichtung und Wahlfreiheit. Konsens bestand darin, allgemeine Prinzipien auf Hochschulebene mit fachspezifischen Regelungen zu verbinden und zugleich Wahlfreiheit zu sichern, sodass auch Nichtnutzende nicht benachteiligt werden.

Erste Transferergebnisse aus dem Blitzlicht

Im abschließenden Blitzlicht nannten Teilnehmende konkrete Vorhaben für ihre Hochschule, wie Kurzleitfäden für Gespräche über KI, Evaluationszyklen für Regelungen, Peer-Programme als niedrigschwellige Qualifizierungsangebote, oder Ringvorlesungen als sichtbare Startpunkte für systematische Qualifizierung.

Grenzen und Möglichkeiten des Formats

Die Ergebnisse spiegeln die Perspektiven der Anwesenden wider und sollten durch einen breiteren Einsatz des Formats ergänzt werden, um die Repräsentativität zu erhöhen. Die Ergebnisse spiegeln die Perspektiven der Workshop-Gebenden und Anwesenden wider und sollten durch einen breiteren Einsatz des Formats ergänzt werden, um die Repräsentativität zu erhöhen. Denkbar wäre zum Beispiel eine Fortführung und Weiterentwicklung des Formates in folgenden Kohorten der DigitalChangeMaker-Initiative, etwa im Rahmen von entsprechenden Veranstaltungen des HFD oder auf einschlägigen Tagungen zum Einsatz von KI in der Hochschullehre. Das Zeitfenster von 90–120 Minuten erwies sich als ausreichend, um Probleme sichtbar zu machen und erste Lösungen zu skizzieren, nicht jedoch für verbindliche Vereinbarungen. Dafür braucht es Folgeprozesse, Termine innerhalb einer zuständigen Gruppe oder feste Evaluationszyklen. Zudem adressieren viele Ideen strukturelle Veränderungen, die ohne klare Zuständigkeiten und Ressourcen Gefahr laufen, zu versanden. Im Workshop wurde außerdem deutlich, dass solche Beteiligungsformate nur dann langfristig Wirkung entfalten können, wenn sie durch geeignete Rahmenbedingungen gestützt werden. Genannt wurden dabei u.a. ihre institutionelle Verankerung in Entscheidungsprozessen und Curricula, die Anerkennung studentischer Expertise sowie die Bereitstellung zeitlicher, finanzieller und organisatorischer Ressourcen.

Ausblick & Fazit

Der Workshop hat gezeigt, dass ein strukturierter Dialog zwischen Studierenden, Lehrenden und Entscheider:innen nicht nur neue Ideen hervorbringt, sondern vor allem gegenseitiges Verständnis schafft. Viele Beteiligte gaben an, studentische Perspektiven bislang kaum gekannt zu haben. Für die Studierenden wiederum war es bedeutsam, ihre Erfahrungen einbringen zu können und ernst genommen zu werden. Als förderlich wurde dabei der geschützte Rahmen des Workshops beschrieben, der trotz der statusgruppenübergreifenden Zusammensetzung durch klare methodische Struktur, moderierte Austauschformate und eine bewusst gleichrangige Gesprächssituation hergestellt wurde. Dies ermöglichte es, auch sensible Themen offen anzusprechen.

Dabei wurden zentrale Spannungsfelder sichtbar: das Verhältnis von Kontrolle und Vertrauen im Umgang mit »KI-sicheren Prüfungen«, die fehlende Transparenz und Einheitlichkeit von Regelungen, ungleiche Voraussetzungen beim Erwerb von Kompetenzen sowie die Frage nach fairer Teilhabe. Zugleich zeigte sich, dass KI Lernprozesse inklusiver gestalten kann, wenn Zugänge barrierearm sind und Studierende selbstbestimmt über den Einsatz entscheiden können.

Deutlich wurde außerdem, dass Partizipation über symbolische Einbindung hinausgehen muss. Statusgruppenübergreifende Formate ermöglichen konstruktive Lösungen und machen die enge Verschränkung der Dimensionen sichtbar: Haltung prägt Regelwerke, Zugänge beeinflussen Qualifizierung, und partizipative Verfahren schaffen Vertrauen und Transparenz. Entscheidend ist eine institutionelle Beziehungskultur, die diese Zusammenhänge anerkennt und Hochschulen befähigt, den KI-gestützten Wandel kooperativ zu gestalten.

Die Ergebnisse fließen über das HFD in den laufenden hochschulpolitischen Diskurs zu KI in der Hochschulbildung ein und verstärken den Anspruch, studentische Stimmen systematisch in hochschulpolitische Aushandlungsprozesse einzubeziehen. Perspektivisch gilt es, Formate wie dieses hochschulintern und in Konferenzkontexten zu etablieren. Die Verbindung von Kollegialer Fallberatung und World Café hat sich dafür als praktikabler und produktiver Ansatz erwiesen.

Die Ergebnisse des Workshops deuten darauf hin, dass partizipative Formate zentrale Bausteine einer fairen und zukunftsfähigen Hochschulentwicklung im Zeitalter von KI sind. Entscheidend ist dabei, ob Hochschulen Strukturen entwickeln, die Studierende dauerhaft als Mitgestaltende anerkennen. Davon hängt ab, ob der technologische Wandel durch KI als Chance genutzt werden kann oder bestehende Ungleichheiten und Misstrauenskulturen fortgeschrieben oder sogar verstärkt werden.

Literatur & weiterführende Quellen

Becker, Sarah/Biehl, Annalisa (2025a): Haltung & Partizipation, Blogbeitrag, Hochschulforum Digitalisierung, https://hochschulforumdigitalisierung.de/haltung-und-partizipation/ vom 29.07.25.

Becker, Sarah/Biehl, Annalisa (2025b): Regelungen im Kontext von Lernen und Prüfen, Blogbeitrag, Hochschulforum Digitalisierung, https://hochschulforumdigitalisierung.de/digitalchangemaker-lernen-und-pruefen/ vom 06.08.25.

Becker, Sarah/Biehl, Annalisa (2025c): Qualifizierung, Blogbeitrag, Hochschulforum Digitalisierung, https://hochschulforumdigitalisierung.de/ki-qualifizierung/ vom 23.10.25.

Biehl, Annalisa (2024): »Schein-Partizipation überwinden – Entwurf eines Partizipationsmodells für Fachschaften«, in: Zeitschrift für Hochschulentwicklung, 19(3), S. 133–151.

Bosse, Elke/Lübcke, Maren (2024): Peer-to-Peer- Strategieberatung – Studie zu Wirkungen, Weiterentwicklung und Transfer des Programms, Arbeitspapier Nr. 82/OKTOBER 2024, Hochschulforum Digitalisierung, https://hochschulforumdigitalisierung.de/wp-content/uploads/2024/10/20241001_HFD_AP_82_P2P-Beratung.pdf

Bovill, Catherine (2020): »Co-creation in learning and teaching: The case for a whole-class approach in higher education«, in: Higher Education, 79(6), S. 1023–1037.

Budde, Jannica/Tobor, Jens (2025): KI-Monitor 2025: Hochschulen gestalten den KI-Alltag, Hochschulforum Digitalisierung/CHE, https://hochschulforumdigitalisierung.de/wp-content/uploads/2025/09/Blickpunkt_KI-Monitor25.pdf

Dietrich, Sieke/Ditzel, Benjamin/Rasch, Sabine/Schrader, Saskia (2025): Editorial: »Studentische Partizipation«, in: Horizont Lehre. Perspektiven und Wege, 1(1), August.

Ditzel, Benjamin (2025): »Studentische Partizipation: Eine demokratie-, organisations- und lerntheoretische Standortbestimmung«, in: Horizont Lehre. Perspektiven und Wege, 1(1), August.

Gostmann, Inga (2025): Digitale Teilhabe & Zugänge, Blogbeitrag, Hochschulforum Digitalisierung, https://hochschulforumdigitalisierung.de/ki-digitale-teilhabe-und-zugaenge/ vom 07.10.25.

Hadar Shoval, Dorit (2025): »Artificial Intelligence in Higher Education: Bridging or Wideningn the Gap for Diverse Student Populations?«, in: Education Science, 15 (5), 637, https://doi.org/10.3390/educsci15050637

Hochschulforum Digitalisierung (2025a): Die Methode: Kollegiale Fallberatung, https://hochschulforumdigitalisierung.de/wp-content/uploads/2025/01/Die-Methode-Kollegiale-Fallberatung.pdf

Hochschulforum Digitalisierung (o.D.): DigitalChangeMaker x KI-Sprint, https://hochschulforumdigitalisierung.de/dcm-ki-sprint-studentische-forderungen-ki-an-hochschulen/

Hüsch, Marc/Horstmann, Nina/Breiter, Andreas (2024): CHECK – Künstliche Intelligenz in Studium und Lehre: Die Sicht der Studierenden im Wintersemester 2023/24. Gütersloh: CHE.

Hüsch, Marc/Horstmann, Nina/Breiter, Andreas (2025): DatenCHECK 6/2025: Künstliche Intelligenz im Studium – die Sicht von Studierenden im Wintersemester 2024/25. Gütersloh: CHE. https://hochschuldaten.che.de/kuenstliche-intelligenz-im-studium-die-sicht-von-studierenden-im-wintersemester-2024-25/

Die KI-Experimentierumgebung an der FernUniversität in Hagen

Erste Erfahrungen aus der Lehre mit generativer KI[1]

Anna Hinzmann und Ferdal Özcelik

Der Beitrag beschreibt die KI-Experimentierumgebung an der FernUniversität in Hagen sowie erste Erfahrungen aus der Erprobung generativer KI in der Lehre. Die Umgebung integriert verschiedene generative KI-Modelle, darunter Open-Weights-Modelle über ein webbasiertes Interface sowie ChatGPT, das über ein Moodle-Plugin in die Lernplattform eingebunden ist. Die Einführung der Tools wurde durch eine formativ, explorativ angelegte Begleitforschung unterstützt, deren Ziel es war, Nutzungsmuster, Akzeptanz sowie wahrgenommene Potenziale und Herausforderungen im Lehrkontext zu erfassen und Hinweise für die Weiterentwicklung der Infrastruktur abzuleiten. Die bisherigen Beobachtungen aus der einjährigen Pilotphase deuten darauf hin, dass der Betrieb lokaler Large Language Models (LLMs) technisch realisierbar ist, derzeit jedoch mit Blick auf Wartungsaufwand und Skalierbarkeit Grenzen aufweist. Das Moodle-Plugin lässt sich grundsätzlich gut in bestehende Kursumgebungen integrieren, zeigt jedoch funktionale Einschränkungen, etwa fehlende Chathistorien oder individuelle Einstellungsmöglichkeiten. Insgesamt blieb die Nutzung der bereitgestellten Tools eher selektiv; zugleich zeigte sich, dass eine gezielte didaktische Einbettung in konkrete Lernaufgaben die Nutzung deutlich begünstigt.

The AI experimental environment at FernUniversität in Hagen: Initial experiences from teaching with generative AI.

The article describes the AI experimentation environment at the FernUniversität in Hagen and initial experiences from testing generative AI in teaching. The environment integrates various generative AI models, including OpenWeights models via a web-based interface and ChatGPT, which is integrated into the learning platform via a Moodle plugin. The introduction of the tools was supported by formative, exploratory accompanying research, the aim of which was to record usage patterns, acceptance, and perceived potential and challenges in the teaching context and to derive recommendations for the further development of the infrastructure. Observations from the one-year pilot phase to date indicate that the operation

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

of local large language models (LLMs) is technically feasible, but currently has limitations in terms of maintenance and scalability. The Moodle plugin can generally be integrated well into existing course environments, but has functional limitations, such as a lack of chat histories or individual setting options. Overall, the use of the tools provided remained rather selective; at the same time, it became apparent that targeted didactic embedding in specific learning tasks significantly promotes use.

Einleitung

Die rasante Entwicklung von KI-Technologien eröffnet neue Möglichkeiten für die Hochschullehre und wirft zugleich Fragen nach geeigneten didaktischen, technischen und organisatorischen Rahmenbedingungen ihres Einsatzes auf. Um diese Potenziale systematisch zu erproben, haben das Zentrum für Lernen und Innovation (ZLI) und das Forschungszentrum CATALPA (Center of Advanced Technology for Assisted Learning and Predictive Analytics) an der FernUniversität in Hagen im Jahr 2024 eine KI-Experimentierumgebung entwickelt, die Lehrenden und Studierenden einen niedrigschwelligen Zugang zu generativen KI-Tools in der Lehre ermöglicht. Ziel war es, unterschiedliche Einsatzszenarien unter realen Studienbedingungen zu erproben und institutionelle Erfahrungen für eine nachhaltige Integration generativer KI in Lehr- und Lernprozesse zu sammeln.

Die Einführung der Umgebung wird durch eine explorativ angelegte Begleitforschung unterstützt. Diese dient weniger dem Nachweis konkreter Wirkungen als vielmehr der Identifikation von Nutzungsmustern, Akzeptanz, wahrgenommenen Potenzialen sowie bestehenden Herausforderungen im praktischen Einsatz. Auf dieser Grundlage sollen Stärken und Schwachstellen der bereitgestellten Infrastruktur sichtbar gemacht und Ansatzpunkte für ihre Weiterentwicklung abgeleitet werden.

Der Beitrag stellt zunächst Konzeption und technische Ausgestaltung der KI-Experimentierumgebung vor und berichtet anschließend über zentrale Beobachtungen und Erfahrungen aus der Pilotphase. Im Mittelpunkt stehen dabei Nutzungsmuster, Akzeptanz und didaktische Einbettung der eingesetzten KI-Tools sowie daraus abgeleitete Gestaltungs- und Unterstützungsbedarfe für die weitere Implementierung generativer KI in der Hochschullehre.

Die KI-Experimentierumgebung der FernUniversität in Hagen

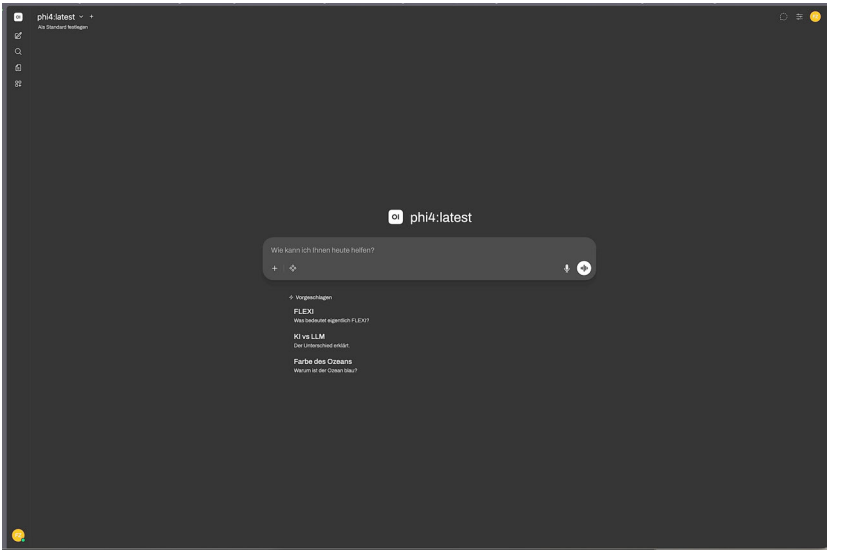
Im Rahmen der KI-Experimentierumgebung wurden an der FernUniversität zwei Szenarien für den Einsatz generativer KI in der Lehre umgesetzt. Der Implementierung ging ein mehrstufiger Auswahl- und Abstimmungsprozess voraus, in dem ins-

besondere rechtliche Rahmenbedingungen (u.a. Datenschutz und Urheberrecht), technische Umsetzbarkeit, Kosten sowie didaktischer Mehrwert systematisch gegeneinander abgewogen wurden. Ziel war es, sowohl experimentelle Freiräume für innovative Nutzungsszenarien zu schaffen als auch eine niedrigschwellige Integration in bestehende Lehr-Lern-Strukturen zu ermöglichen.

Realisiert wurden zwei unterschiedliche Zugänge:

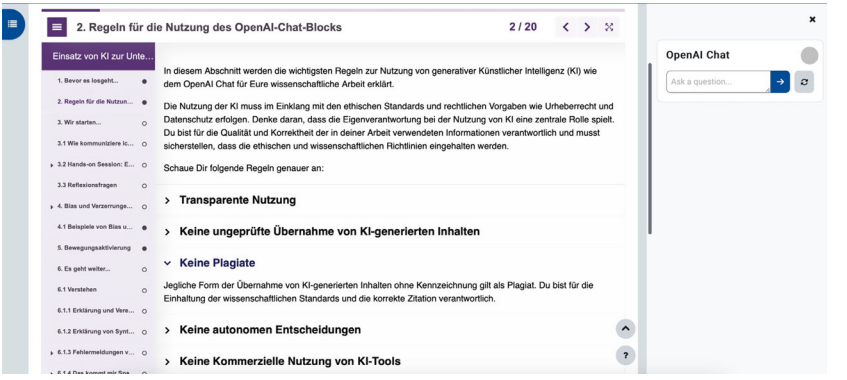
Über ein webbasiertes Interface (OpenWebUI) stellte das Forschungszentrum CATALPA mehrere Open-Weights-LLMs auf der hochschuleigenen Infrastruktur bereit (FLEXI; vgl. Zesch et al., 2024). Dieses Szenario erlaubt eine flexible Anpassung und Erprobung verschiedener Modelle sowie die Verarbeitung sensibler oder urheberrechtlich geschützter Materialien innerhalb der lokalen Systemumgebung.

Abbildung 1: Flexi-Plattform



Ergänzend wurde über ein Plugin die Einbindung von ChatGPT in die Lernplattform Moodle realisiert. Die Anbindung erfolgt über die Microsoft Azure-Cloud und ermöglicht eine datenschutzkonforme Nutzung aktueller GPT-Modelle innerhalb der bestehenden Kursumgebungen.

Abbildung 2: »OpenAI Chat Block«-Plugin in Moodle-Lernumgebung



Beide Ansätze verfolgen unterschiedliche Zielsetzungen und ergänzen sich funktional: Während die lokale Flexi-Plattform vor allem hohe Anpassbarkeit und experimentelle Freiräume bietet, ermöglicht das Moodle-Plugin eine nahtlose Integration in die vertraute Lernumgebung und damit einen besonders niedrigschwelligen Zugang für Studierende. Durch diese Kombination sollte sowohl exploratives Ausprobieren als auch eine breite praktische Nutzung im Lehralltag unterstützt werden.

Start der Pilotphase

Um das Interesse an KI-gestützten Lehrmethoden hochschulweit zu fördern und unterschiedliche Einsatzszenarien unter realen Lehrbedingungen erproben zu können, wurde im Sommer 2024 ein offener Aufruf zur Teilnahme an der Pilotphase gestartet. Lehrende aller Fakultäten waren eingeladen, eigene didaktische Konzepte und Nutzungsideen für generative KI einzureichen.

Insgesamt gingen 30 Projektvorschläge ein, die hinsichtlich rechtlicher, technischer und didaktischer Anforderungen gebündelt und zu fünf Clustern mit jeweils ähnlichen Anwendungsszenarien zusammengefasst wurden. Die Bandbreite reichte von fachspezifischen Chatbots auf Basis von Retrieval-Augmented Generation (RAG) über Unterstützungsangebote beim Programmieren – etwa in R – bis hin zum Einsatz generativer KI beim wissenschaftlichen Arbeiten, bei Textanalysen oder der Prüfungsvorbereitung.

Die beteiligten Lehrenden wurden durch das Zentrum für Lernen und Innovation (ZLI) kontinuierlich beraten und je nach Szenario den geeigneten technischen Lösungen zugeordnet. Neben einer Einführung in die Nutzung der jeweiligen Tools

umfasste die Begleitung auch didaktische Beratung zur sinnvollen Einbettung der Anwendungen in bestehende Lehr-Lern-Settings.

Die Pilotphase wurde zudem durch eine formative Begleitforschung unterstützt. Ziel war es, Nutzungspraxen, Akzeptanz und auftretende Herausforderungen systematisch zu erfassen und die Weiterentwicklung der Experimentierumgebung datenbasiert zu informieren. Hierzu wurden unter anderem standardisierte Kurzbefragungen von Studierenden, Fokusgruppengespräche mit Lehrenden sowie systemseitige Nutzungsdaten (z.B. Tokenverbrauch) herangezogen.

Die Nutzung der KI-Experimentierumgebung

Im Verlauf der Pilotphase zeigte sich eine differenzierte Nutzung der beiden bereitgestellten Zugänge, die sich jeweils in unterschiedlichen didaktischen Kontexten bewährten.

Die lokal betriebene FLEXI-Plattform wurde insbesondere für experimentelle und fachspezifische Anwendungen genutzt. Schwerpunkte lagen vor allem in der Unterstützung bei Programmieraufgaben, im wissenschaftlichen Arbeiten sowie im Vergleich und Test unterschiedlicher Open-Weights-LLMs. Die Plattform diente damit vor allem als explorative Entwicklungs- und Erprobungsumgebung, in der Lehrende und Studierende verschiedene Modelle und Konfigurationen flexibel ausprobieren konnten. Aufwendigere Szenarien wie der Aufbau von RAG-Systemen erwiesen sich im Projektverlauf jedoch als ressourcenintensiv und konnten bislang nicht umgesetzt werden.

Das über Moodle integrierte Plugin »OpenAI Chat Block« mit ChatGPT wurde demgegenüber breiter im regulären Lehrbetrieb eingesetzt. Im Wintersemester 2024/25 war es in sechs, im Sommersemester 2025 bereits in zwölf Kursumgebungen eingebunden. Die Nutzung umfasste vielfältige studienbezogene Unterstützungsfunktionen, darunter Hilfe bei Programmieraufgaben (z.B. in R), Unterstützung beim wissenschaftlichen Schreiben und bei Abschlussarbeiten, Formulierungs- und Strukturierungshilfen für Texte, Übersetzungen, die Aufbereitung englischsprachiger Fachliteratur sowie analytische und kreative Aufgabenformate, etwa in der Literaturwissenschaft. Darüber hinaus wurde das Tool gezielt eingesetzt, um Möglichkeiten und Grenzen generativer KI gemeinsam mit Studierenden zu reflektieren und kritisch zu erproben.

Insgesamt zeigt sich damit ein komplementäres Nutzungsmuster: Während FLEXI vor allem als flexible Test- und Entwicklungsumgebung diente, ermöglichte das Moodle-Plugin eine niedrigschwellige Integration generativer KI in den regulären Kursalltag.

Erfahrungen aus der Pilotphase mit der KI-Experimentierumgebung

Die Pilotphase lieferte differenzierte Erkenntnisse zur technischen Umsetzbarkeit, Skalierbarkeit und didaktischen Nutzung der beiden Szenarien der KI-Experimentierumgebung.

Hinsichtlich der lokal betriebenen Open-Weights-Modelle zeigte sich, dass deren Bereitstellung grundsätzlich möglich ist, sofern entsprechendes technisches Know-how und geeignete Infrastruktur vorhanden sind. Gleichzeitig wurde deutlich, dass der damit verbundene Betriebs-, Wartungs- und Betreuungsaufwand eine Skalierung und einen Einsatz im Regelbetrieb derzeit begrenzt. Als Vorteil dieses Szenarios erwies sich insbesondere die hohe Flexibilität: Unterschiedliche Modelle können angebunden, individuell konfiguriert und hinsichtlich ihrer Ergebnisse vergleichend erprobt werden. Aufwendigere Anwendungsszenarien wie der Aufbau von RAG-Systemen stellten sich im Projektverlauf jedoch als ressourcenintensiv heraus und konnten im Rahmen der Pilotphase nicht realisiert werden.

Das in die Lernplattform Moodle integrierte Plugin, das über die Microsoft Azure-Infrastruktur datenschutzkonform an ChatGPT angebunden ist, erwies sich demgegenüber als technisch stabil und gut skalierbar. Die Qualität der generierten Antworten hängt dabei wesentlich vom jeweils angebundenen Modell ab, ebenso wie die entstehenden Token-Kosten. Als funktionale Einschränkungen wurden insbesondere das Fehlen einer Chathistorie, eingeschränkte Möglichkeiten zur individuellen Konfiguration (z.B. individuelle Temperatureinstellungen, Systemprompts etc.) sowie die begrenzte Größe des Ein- und Ausgabefensters identifiziert.

Aus didaktischer Perspektive zeigte sich, dass die Nutzung der bereitgestellten KI-Tools im Rahmen der Pilotphase selektiv blieb. Studierende, die das Moodle-Plugin nutzten, setzten es vor allem zur Klärung von Verständnisfragen, zur Textarbeit, zur Prüfungsvorbereitung sowie zur Unterstützung bei der Entwicklung von Konzepten ein. Die Nutzung wurde insgesamt als zumindest teilweise hilfreich wahrgenommen; vereinzelt berichteten Studierende zudem von einer leichten Unterstützung ihrer Motivation sowie ihrer Fähigkeit zur eigenständigen Erarbeitung von Inhalten.

Deutlich wurde zudem, dass die didaktische Einbettung einen zentralen Einfluss auf die tatsächliche Nutzung hat. Insbesondere in Kursen, in denen das KI-Tool explizit mit konkreten Übungs- oder Arbeitsaufgaben verknüpft war, wurde es häufiger eingesetzt. Aus Sicht der Studierenden besteht ein Bedarf an stärker kontextualisierten Angeboten, etwa in Form studiengangs- oder aufgabenspezifischer Chatbots oder klar definierter Einsatzszenarien für generative KI.

Übergreifend lässt sich ein hoher Unterstützungs- und Qualifizierungsbedarf sowohl bei Lehrenden als auch bei Studierenden feststellen. Während einzelne Leh-

rende und Studierende bereits über umfangreiche Erfahrung im Umgang mit generativer KI verfügen, besteht bei vielen anderen noch erheblicher Orientierungs- und Schulungsbedarf.

Ausblick

Die Pilotierung des in Moodle integrierten KI-Plugins wird zunächst fortgeführt, um weitere Nutzungserfahrungen in unterschiedlichen Lehrkontexten zu sammeln und die didaktische Einbettung weiter zu erproben. Parallel werden aktuelle Entwicklungen im Moodle-Ökosystem sowie neue Plugins und Anbindungsmöglichkeiten für generative KI systematisch beobachtet und hinsichtlich ihrer Eignung getestet, da das derzeit eingesetzte »OpenAI Chat Block«-Plugin funktional nur eingeschränkte Konfigurations- und Interaktionsmöglichkeiten bietet.

Ergänzend wurde mit KI:connect[2] an der FernUniversität in Hagen ein eigenständiger, plattformunabhängiger Zugang zu generativer KI geschaffen, der seit September 2025 auch Studierenden zur Verfügung steht. Damit entsteht eine zusätzliche Option zur Nutzung generativer KI außerhalb der Lernplattform, die im Zusammenspiel mit den bestehenden Angeboten explorativ erprobt und vergleichend begleitet werden soll.

Ein weiterer Schwerpunkt liegt auf dem Ausbau von Unterstützungs- und Qualifizierungsangeboten. Geplant sind unter anderem Schulungen für Lehrende sowie ein studienbegleitendes KI-Modul für Studierende, um Kompetenzen im reflektierten und fachlich angemessenen Einsatz generativer KI systematisch zu fördern.

Perspektivisch wird zudem die Entwicklung stärker kontextualisierter und personalisierter Assistenzsysteme, etwa kurs- oder fachbezogener Chatbots, angestrebt. Solche Lösungen könnten insbesondere im Fernstudium niedrigschwellige Unterstützung bieten und individuelle Lernprozesse begleiten.

Die genannten Maßnahmen werden weiterhin durch eine formative Begleitung unterstützt, um Nutzungspraxen, Bedarfe und Herausforderungen fortlaufend zu erfassen und die bereitgestellten Angebote schrittweise an die Anforderungen von Lehre und Studium anzupassen.

2 KI:connect.nrw ist das landesweite Portal für generative KI-Anwendungen an Hochschulen in Nordrhein-Westfalen, das Hochschulangehörigen den Zugang zu verschiedenen KI-Modellen ermöglicht.

Literatur

Zesch, Torsten/Hanses, Michael/Seidel, Niels/Aggarwal, Piush/Veiel, Dirk/de Witt, Claudia (2024): FernUni LLM Experimental Infrastructure (FLEXI) – Enabling Experimentation and Innovation in Higher Education Through Access to Open Large Language Models. https://doi.org/10.48550/arXiv.2407.13013 vom 27.06.2024.

Implementierung, Nutzung und Herausforderungen eines hochschuleigenen KI-Interfaces[1]

Isabella Buck[2], Nina Christ, Alexander Kaib und Lisa Ulzheimer

Dieser Beitrag stellt die Entwicklung, Implementierung und Nutzung des hochschuleigenen KI-Interfaces der Hochschule RheinMain vor. Über ein zentrales Web-Frontend werden hier verschiedene KI-Dienste wie Textgenerierung, Transkription, Bildgenerierung und Übersetzung integriert. Mittels einer Online-Befragung und qualitativer Interviews wurde die Nutzung des Interfaces im SoSe 2025 evaluiert. Die Ergebnisse zeigen, dass Lehrende das hochschuleigene Angebot häufiger nutzen als Studierende, primär aufgrund von Datenschutzaspekten. Studierende bevorzugen hingegen oft kommerzielle Tools wegen besserer Benutzerfreundlichkeit und der Vertrautheit damit. Ein zentrales Problem ist die mangelnde Sichtbarkeit und Kommunikation der Vorteile von KI@HSRM, insbesondere des didaktisch aufbereiteten Vorlagenbereichs. Die bloße Bereitstellung einer hochschuleigenen Lösung garantiert daher noch keine breite Nutzung.

Implementation, usage and challenges of a custom AI interface built for our university

This paper presents the development, implementation, and usage of the KI@HSRM AI interface at RheinMain University of Applied Sciences. This central web frontend integrates various AI services, including text and image generation, transcription, and translation. An evaluation of the usage of the interface in the summer semester of 2025 was conducted via an online survey and qualitative interviews. The results indicate that faculty use the university-owned platform more frequently than students, primarily due to data privacy concerns. In contrast, students often favor commercial tools, citing better usability and familiarity. A key challenge identified is the lack of visibility and communication regarding the benefits of KI@HSRM, particularly its collection of didactically curated templates. The study concludes that simply providing a university-owned solution does not guarantee widespread adoption.

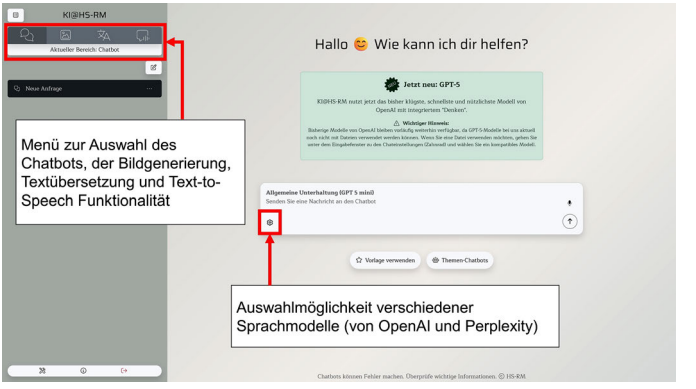
1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000–0003-4222-0426

Einleitung

Viele Hochschulen ermöglichen ihren Angehörigen mittlerweile Zugriff auf KI-Dienste, oft in Form eigener Chat-Interfaces zu verschiedenen Sprachmodellen (vgl. Paaßen et al. 2025). Auch die Hochschule RheinMain stellt unter dem Titel KI@HSRM eine solche Plattform bereit, die im Sommersemester 2025 erstmals evaluiert wurde. Dieser Beitrag stellt die Ergebnisse der durchgeführten Online-Befragungen und Interviews vor, um Aufschluss über Akzeptanzfaktoren und Nutzungsmuster hochschuleigener KI-Lösungen zu geben, insbesondere im Vergleich zu kommerziellen Alternativen. Zur Einordnung von KI@HSRM gehen wir zunächst kurz auf den Entstehungskontext und Funktionsumfang der Plattform ein.

Abbildung 1: Screenshot von KI@HSRM



KI@HSRM ist eingebettet in eine umfassende, institutionell verankerte Auseinandersetzung mit generativer KI, die Anfang 2023 begann. In einem partizipativen Prozess wurde u.a. ein hochschulweiter Think Tank etabliert, in dem Hochschulangehörige gemeinsam Perspektiven für eine zukunftsfähige Hochschulbildung unter Bedingungen von KI erarbeiteten (vgl. Buck et al. 2023). Im Februar 2024 beschloss das Präsidium die modellhafte Einführung einer auf der OpenAI-API basierenden KI-Plattform, die schrittweise ausgerollt und agil weiterentwickelt wird. Mittlerweile stehen dort auch DeepLs Übersetzungs- und ›Write‹-Funktion, Text-to-Speech sowie Perplexity Sonar zur Verfügung. Perspektivisch ist eine Erweiterung um zusätzliche Komponenten wie lokal betriebene LLMs vorgesehen. Ein zentrales Merkmal des Projekts ist die didaktische Einbettung der technischen Infrastruktur u.a. durch einen Vorlagenbereich mit Prompts für Studierende und Lehrende. Schulungen, Workshops und ein KI-Newsletter ergänzen das Angebot.

Die Plattform ist damit Teil eines umfassenden hochschuldidaktischen Transformationsprozesses und unterstützt Lehr- und Lernprozesse unter Berücksichtigung hochschulspezifischer Bedarfe.

Ergebnisse der Nutzendenbefragung

Die Evaluation des hochschuleigenen KI-Interfaces erfolgte a) mittels eines Online-Fragebogens, den von Ende Mai bis Mitte Juli 2025 74 Studierende und 78 Lehrende ausfüllten, sowie b) mittels qualitativer Interviews mit sechs Lehrenden und zwei Studierenden von April bis Juni 2025. Erreicht wurden die Befragungsteilnehmenden über einen hochschulinternen Newsletter, Webseiten und ein Pop-Up-Fenster im KI-Interface. Es wurden gezielt nur Personen angesprochen, die KI@HSRM bereits benutzt hatten, entsprechend beziehen sich unsere Ergebnisse auf diese Gruppe. Die Interviews dauerten von 10 bis 18 Minuten; sie wurden qualitativ inhaltsanalytisch ausgewertet (Mayring 2022). Aus dem Interviewleitfaden wurden deduktiv die vier Kategorien »Anwendungsfälle und Akzeptanz«, »Interaktion und Benutzerfreundlichkeit«, »Herausforderungen und Verbesserungsmöglichkeiten« sowie »Schlussfolgerungen/Empfehlungen« herausgearbeitet.

Ergebnisse der Studierendenbefragung

Von den befragten Studierenden (70 % im Bachelor-, 24 % im Master- und 6 % im Promotionsstudium; n=67) gaben 41 % an, das KI-Interface weniger als monatlich zu nutzen, gefolgt von 37 %, die es wöchentlich nutzen (n=68). Im Vergleich zu der deutschlandweit (sehr) häufigen Nutzung von KI-Tools unter Studierenden (von Garrel/Mayer 2025: 20) wird KI@HSRM also eher selten verwendet, was wir darauf zurückführen, dass ein Großteil der Befragten (79 %, n=68) parallel auch kommerzielle KI-Tools einsetzt.

Die wichtigsten Vorteile kommerzieller Tools gegenüber dem Hochschul-Interface waren laut den Studierenden die höhere Benutzerfreundlichkeit (51 %), die Funktionen und Features (49 %) sowie die Ergebnisqualität (49 %) (n=45). Erfahrungen mit kommerziellen Tools beeinflussen die Entscheidung, KI@HSRM zu nutzen, bei 70 % der Befragten stark oder sehr stark (n=43). So gab mit 37 % die größte Gruppe an, künftig vorzugsweise kommerzielle Produkte zu nutzen, während 35 % angaben, die Nutzung beizubehalten bzw. 28 % KI@HSRM häufiger zu verwenden (n=43).

Die Zufriedenheit mit dem hochschuleigenen Interface ist gespalten: Je 38 % der Befragten sind zufrieden oder sehr zufrieden bzw. unzufrieden oder sehr unzufrieden; 24 % positionierten sich neutral (n=63). Als primäre Anwendungsfälle wurde Lernen bzw. die Klausurvorbereitung (49 %; n=61) sowie die Ideenentwicklung und

sprachliche Verbesserung von Texten (je 46 %) genannt. Der didaktisch konzipierte Vorlagenbereich wird dabei von der Mehrheit der Nutzenden (62,7 %; n=59) nie verwendet, was primär auf fehlendes Wissen über dessen Existenz zurückgeführt wird: Die meisten Studierenden, die die Vorlagen nicht benutzen, kannten sie nicht (56,8 %).

Zur vertiefenden Einordnung der quantitativen Ergebnisse sollten leitfadengestützte Interviews geführt werden. Trotz mehrfacher Anfragen über diverse Kanäle erklärten sich am Ende lediglich zwei Studierende dazu bereit. Auch wenn die geringe Zahl keine generalisierbaren Aussagen erlaubt, lassen sich erste Hinweise auf Akzeptanzfaktoren identifizieren. Als zentrale Gründe für die Nutzung des hochschuleigenen Interfaces wurden vor allem Datenschutzaspekte und die Kostenfreiheit genannt. Die Interviewten schätzten, dass bei KI@HSRM im Unterschied zu ChatGPT keine Tokenlimits bestünden und sich beliebig viele Dateien hochladen lassen.

Trotz dieser positiven Aspekte griffen beide Interviewten auch regelmäßig auf kommerzielle Tools zurück. ChatGPT wurde dabei vor allem für informelle, nichtakademische Zwecke genutzt, etwa beim Kochen. Als ausschlaggebend für diese Präferenz wurden die bessere Usability, insbesondere über die mobile App, die für KI@HSRM nicht existiert, genannt.

Beide Studierenden machten konkrete Vorschläge zur Steigerung der Sichtbarkeit und Nutzung von KI@HSRM. Wiederholt wurde der Wunsch geäußert, die Plattform verstärkt in der Studieneingangsphase zu bewerben. Während die eine Person gezielte Mailings für sinnvoll hielt, erachtete die andere diese als nicht zielführend und schlug vor, dass Lehrende in Veranstaltungen aktiv auf das Angebot hinweisen sollten. Insgesamt wurde ein Mangel an Sichtbarkeit konstatiert – sowohl bzgl. des Interfaces selbst als auch dessen konkreter Funktionen und vor allem bzgl. der Vorteile in Bezug auf kommerzielle Lösungen.

Ergebnisse der Lehrendenbefragung

Bei den befragten Lehrenden zeigt sich eine deutlich höhere Nutzung des KI-Interfaces: 20 % nutzen es täglich und 55,4 % wöchentlich (n=65). Ferner nutzen deutlich weniger Lehrende als Studierende, insgesamt 62 % der Lehrenden (n=65), parallel auch kommerzielle Tools. Von denen, die auch solche nutzen, setzen 85 % diese in ähnlichen Kontexten wie KI@HSRM ein (n=34). Als Hauptvorteile kommerzieller Tools wurden die Ergebnisqualität (41 %), die Funktionen (35 %) und die Benutzerfreundlichkeit (32 %) genannt. Nur 9 % gaben an, keine Vorteile zu sehen. Im Vergleich zu den Studierenden beeinflussen kommerzielle Tools die Nutzung von KI@HSRM bei Lehrenden geringer: Nur 39 % geben an, von diesen stark oder sehr stark beeinflusst zu sein, 33 % mäßig und 27 % gering oder überhaupt nicht (n=33). Im Gegensatz zu den Studierenden planen die Lehrenden eher eine gleichbleibende

Nutzung von KI@HSRM (55 %); nur eine Minderheit (15 %) will künftig v.a. kommerzielle Produkte verwenden (n=33).

Die Zufriedenheit mit KI@HSRM ist bei den Lehrenden höher als bei den Studierenden: 58 % sind zufrieden oder sehr zufrieden, während nur 10 % unzufrieden bis sehr unzufrieden sind (n=60). Die am häufigsten genannten Anwendungsszenarien sind die Erstellung von Lehrmaterialien und Aufgabenstellungen (je 49 %) und das Erledigen von Verwaltungsaufgaben (40 %) (n=57). Auch bei den Lehrenden wurden die spezifischen Vorlagen kaum genutzt (51 % »nie«, 35 % »selten« (n=55)), was ebenfalls primär auf fehlendes Wissen über deren Existenz zurückgeführt wird (71,4 %).

Die parallel zur quantitativen Befragung durchgeführten qualitativen Interviews mit sechs Lehrenden liefern differenziertere Einblicke in Gründe für die Nutzung von KI@HSRM und wahrgenommene Hindernisse für einen breiteren Einsatz des hochschuleigenen Interfaces.

Als primärer bzw. teilweise sogar einziger Grund für den Einsatz von KI@HSRM statt kommerzieller Alternativen wurde Datenschutz bzw. Datenhoheit angeführt (»Datenschutz. Das ist der ganze Grund« (LEHR2)). Des Weiteren wurde die Verlässlichkeit des Interfaces im Hinblick auf die Stabilität und Verfügbarkeit hervorgehoben sowie angemerkt, dass ein Nutzungsgrund in der Möglichkeit liege, eigene Dateien hochzuladen. Auch die Bequemlichkeit eines zentralen Zugangs ohne zusätzliche Accounts wurde genannt. Zwei Lehrende gaben an, seit dem Launch von KI@HSRM nur noch dieses KI-Tool zu nutzen. Die anderen vier befragten Lehrenden greifen weiterhin auch auf kommerzielle Tools zurück, unter anderem um die Outputs verschiedener Tools abzugleichen.

Als Hemmnis für eine breitere Nutzung von KI@HSRM unter Studierenden wurde von mehreren Lehrenden ein fehlendes Wissen über das Angebot vermutet. Einige gaben an, ihre eigenen Studierenden würden das Tool regelmäßig einsetzen, während andere keinerlei Rückmeldung aus der Studierendenschaft erhielten. Der Eindruck, dass viele Studierende bereits mit anderen Tools vertraut seien und daher keinen Anlass zum Umstieg sähen, wurde mehrfach geäußert. In Bezug auf Kolleg:innen berichtete eine interviewte Person eine ähnliche Tendenz zu kommerziellen Tools, wobei andere Befragte ein hohes Interesse an KI@HSRM wahrnahmen.

Zur Erhöhung der Sichtbarkeit und Nutzung des hochschuleigenen Interfaces wurden vielfältige Vorschläge gemacht. Genannt wurden u.a. eine bessere Bewerbung der Vorteile von KI@HSRM gegenüber kommerziellen Lösungen sowie deren Zusatzfunktionen wie den Vorlagenbereich, eine Werbekampagne in Erstsemesterveranstaltungen, ein prägnanterer Name, eine bessere Auffindbarkeit auf der Hochschul-Website sowie die Entwicklung einer mobilen/desktopbasierten Version. Als zentral wurde, ebenso wie in den beiden Interviews mit Studierenden, hervorgehoben, dass Lehrende eine Vorbildfunktion einnehmen: Eine gezielte Informations-

kampagne der Lehrenden über Dekanate oder in Fachbereichssitzungen könne dazu beitragen, das Interface verstärkt in die Lehre zu integrieren, um es so auch unter Studierenden bekannter zu machen.

Diskussion

Die Evaluation von KI@HSRM zeigt, dass die Bereitstellung eines hochschuleigenen Interfaces trotz regelmäßiger, zielgruppenspezifischer Schulungsangebote keine breite Nutzung garantiert, besonders unter Studierenden. Dafür sprechen sowohl die geringen Teilnahmezahlen dieser Zielgruppe sowie ihr gespaltenes Antwortverhalten: Akzeptanz und Ablehnung finden sich hier in gleichem Maße. Im Gegensatz dazu sind Lehrende mit der hochschuleigenen Lösung insgesamt zufrieden und nutzen sie häufiger.

Auch wenn unsere Datenbasis keine abschließende Erklärung dieser Tendenzen zulässt, lassen sich einige Vermutungen aufstellen: Der Einfluss und die Nutzung kommerzieller Tools ist unter Studierenden höher, was darauf schließen lässt, dass diese Gruppe das hochschuleigene Interface eher an den Standards der bekannten Chatbots misst – und mit diesen kann es hinsichtlich Benutzerfreundlichkeit sicher nicht mithalten. Erstaunlich ist, dass knapp die Hälfte der Befragten den kommerziellen Tools eine höhere Antwortqualität zuschreibt, obwohl im Hintergrund dieselben Sprachmodelle verwendet werden. Möglicherweise besteht also auch ein unzureichendes Verständnis über die Funktionsweise generativer KI-Anwendungen. Weiterhin ist es plausibel, dass Studierende im Vergleich zu Lehrenden einen geringeren Druck verspüren, institutionell vorgegebene Tools zu nutzen.

Unsere Evaluation wirft damit eine grundlegende Frage auf, die sich mittlerweile einige Hochschulen stellen dürften: Was ist für Studierende der Mehrwert einer hochschuleigenen Schnittstelle zu LLMs, die sie bereits über benutzerfreundlichere Webanwendungen nutzen? Datenschutz und Kostenersparnis scheinen aktuell nicht genug, um die Mehrheit der Studierenden zu überzeugen – was nicht bedeutet, dass diese Argumente gar nicht wahrgenommen werden.

Vielleicht sind unsere Ergebnisse ein Hinweis darauf, dass die Nachbildung bekannter Chat-Interfaces nicht der beste Weg ist, um die potenziellen Vorteile einer hochschuleigenen KI-Lösung zu implementieren und kommunizieren. Besonders die ernüchternde Unbekanntheit der didaktischen Vorlagen zeigt, dass die Nutzeroberfläche von KI@HSRM zu bekannten Verhaltensmustern einlädt, statt gezielt pädagogisch sinnvolle KI-Interaktionen anzuregen. Sicherlich ist es sinnvoll, Hochschulangehörigen einen niedrigschwelligen Zugang zu Sprachmodellen zur Verfügung zu stellen – was sich nicht zuletzt in den positiven Ergebnissen der Lehrendenbefragung zeigt – doch möglicherweise braucht es auch Funktionen, die sich *nicht* in ChatGPT und Co. wiederfinden, um die breitere Studierendenschaft anzuspre-

chen. So interpretiert stellen unsere Umfrageergebnisse ein weiteres Argument für die Entwicklung und Integration offener KI-Bildungstechnologien (vgl. Paaßen et al. 2025) dar.

Literatur

Buck, Isabella/Jost, Christiane/Kreis-Hoyer, Petra/Limburg, Anika (2023): KI-induzierte Transformation an Hochschulen. Diskussionspapier Hochschulforum Digitalisierung 26. https://hochschulforumdigitalisierung.de/wp-content/uploads/2023/11/HSRM-Diskussionspapier-Nr.-26-KI-induzierte-Transformation-an-Hochschulen-1.pdf

Mayring, Philipp (2022): Qualitative Inhaltsanalyse, Grundlagen und Techniken. 13., überarbeitete Auflage. Weinheim/Basel: Beltz.

Paaßen, Benjamin/Go, Stefanie/Mayer, Maximilian/Kiesewetter, Benjamin/Krüger, Anne/Leschke, Jonas/Stracke, Christian M. (2025): KI-Infrastruktur für Digitale Autonomie an Hochschulen. Forschungsnetzwerk Artificial Intelligence and Digital Autonomy in Research and Education (AIDARE). https://aidare.org/ki-infrastruktur-fur-digitale-autonomie-an-hochschulen/ vom 06.09.2025.

von Garrel, Joerg/Mayer, Jana (2025): Künstliche Intelligenz im Studium – Eine quantitative Längsschnittstudie zur Nutzung KI-basierter Tools durch Studierende. Hochschule Darmstadt. https://doi.org/10.48444/H_DOCS-PUB-533

KI-Tutoren in der Lehre: rechtlich sicher, didaktisch wertvoll, technisch innovativ[1]

Alexander Gerber[2], Lars Lorenz[3] und Ulrich Pfeiffer[4]

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[t] 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[t] 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 niederschwelligen 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[t] 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[t] 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[t] 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[t] 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[t] 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[t] 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[t] 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./Krathwohl, 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.

Offene Künstliche Intelligenz als Chance für Bildungsprozesse[1]

Beatrix Busse[2], Ingo Kleiber[3] und Kathrin Andree[4]

Offene Künstliche Intelligenz (KI) bietet im Vergleich zu proprietären Angeboten und KI-Systemen neben Vorteilen hinsichtlich der Kosten, Sicherheit oder Souveränität eigene Chancen für Bildungsprozesse. Der Beitrag zeigt, dass offene KI nicht nur technische, sondern auch kulturelle und gesellschaftliche Dimensionen umfasst und zu einem modernen, offenen Verständnis von Bildung und Wissenschaft passt. Offene KI erlaubt es, passgenaue und maßgeschneiderte Systeme für die Bildung zu schaffen, fügt sich in ein modernes Bildungs- und Forschungsverständnis ein und bietet Potenziale für den Aufbau von KI-bezogenen aber auch fachlichen Kompetenzen. Der Beitrag beleuchtet diese drei Ansatzpunkte und zeigt, weshalb offene KI relevant für die Lehre, das Lernen und die Forschung im 21. Jahrhundert ist.

Open Artificial Intelligence as an Opportunity for Educational Processes
Compared to proprietary offerings and AI systems, open Artificial Intelligence (AI) offers not only advantages in terms of cost, security, or sovereignty, but also unique opportunities for educational processes. This article discusses how open AI encompasses not only technical but also cultural and social dimensions. Open AI makes it possible to create tailor-made systems for education, fits in with a modern understanding of both education and research, and offers the potential for building AI-related as well as subject-specific competencies. This article highlights these three starting points and demonstrates why open AI is relevant for teaching, learning, and research in the 21st century.

Einleitung

Die Wahrnehmung von (generativer) Künstlicher Intelligenz (KI), ob in der Öffentlichkeit oder im Bildungskontext, ist stark durch die Produkte großer, kommer-

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000-0003-4696-6727
3 ORCID: 0000–0001-7231-2714
4 ORCID: 0009–0006-4613-9966

zieller Anbieter*innen geprägt. Für viele Menschen ist OpenAIs ChatGPT nahezu gleichbedeutend mit generativer KI. Diese Verengung des Blicks auf proprietäre Systeme lenkt davon ab, dass es zunehmend offene, transparente und veränderbare KI-Systeme gibt, die von einem starken Ökosystem getragen werden und zunehmend an Leistungsfähigkeit gewinnen. Mit dieser Offenheit gehen Chancen für die digitale Souveränität, aber auch die Didaktik, einher (vgl. Busse/Kleiber 2025: 175).

Diese didaktischen und auf Bildungsprozesse bezogenen Chancen nehmen bisweilen eine untergeordnete Rolle in den Diskussionen um KI ein. Stattdessen werden offene und proprietäre KI vor allem vor dem Hintergrund der Kosten, der Komplexität, der Leistungsfähigkeit, der Compliance, der Sicherheit, der *Equity* und zunehmend auch der Souveränität diskutiert (vgl. Eiras et al. 2024, Machado 2025, Vake et al. 2025).

Der Beitrag möchte das Zusammenspiel zwischen offener KI und Didaktik näher betrachten. Offene KI eröffnet Chancen für die Bildung, die über die bloße Substitution proprietärer Systeme hinausgehen. Sie ermöglicht (1) passgenaue Lösungen, (2) fügt sich in ein modernes, offenes Bildungs- und Forschungsverständnis ein und (3) schafft Räume für Kompetenzaufbau. Dabei sehen wir offene KI als integratives Konzept, das neben technischen auch kulturelle und gesellschaftliche Dimensionen umfasst.

Was ist offene KI?

Der Begriff »offene KI« wird zumeist technisch verstanden: Sind die Modelle, Daten, Methoden, Algorithmen, Software oder auch die Hardware offen, nachvollziehbar, veränderbar etc.? Dabei ist zu betonen, dass es sich auch bei einer ausschließlich technischen Betrachtung um keine binäre Kategorie handelt. Dies beginnt bei der Differenzierung zwischen Open-Weight- und Open-Source-Modellen und endet bei der Frage danach, ob es auch offene Hardware für echte offene KI braucht.

Wir möchten bewusst über die technische Betrachtung hinausgehen. Daher verstehen wir unter offener KI kein technisches Konzept, sondern eines, das Prinzipien von *Open Source*, *Open Data*, *Open Science* und *Open Education* auf das Feld der KI überträgt und explizit auch eine Kultur der Offenheit und Transparenz beinhaltet. Die Prinzipien, die so mit offener KI verknüpft sind, z.B. Kollaboration, Partizipation, Inklusion, Gemeinwohlorientierung und Unabhängigkeit, sind zentrale Bestandteile des Konzepts. Offene KI umfasst daher sowohl technische Artefakte als auch die Praktiken und Diskurse, die ihre Entwicklung und Nutzung prägen.

Bei alledem sind wir uns aber der zentralen Bedeutung der Technologie bewusst. Daher schließen wir uns Benjamin Paaßen (2025) an, der festhält, dass »offene Sprachmodelle als Technik überhaupt erst die Voraussetzung für digitale Autonomie« bilden.

Offene KI in Bildungskontexten

Wie bereits oben dargestellt, werden Entscheidungen rund um offene KI weiterhin v.a. ausgehend von Faktoren getroffen, die nicht direkt Bildungsprozesse betreffen. Im Folgenden möchten wir daher drei Ansatzpunkte beleuchten: die Entwicklung passgenauer, maßgeschneiderter Systeme, die Passung zum Bildungs- und Forschungsverständnis sowie der Beitrag offener KI zum Kompetenzaufbau.

Passgenaue, maßgeschneiderte KI-Systeme

Eine erste zentrale Chance offener KI liegt in der Möglichkeit, passgenaue und maßgeschneiderte KI-Systeme für konkrete Bildungskontexte und Institutionen zu entwickeln. Dabei geht es sowohl um technische Anpassungen (z.B. eine spezialisierte Benutzer*innenoberfläche) als auch um inhaltliche Anpassungen (z.B. Finetuning von KI-Modellen). Ausgehend von diesen beiden Hauptkategorien können offene KI-Systeme dann an die Rahmenbedingungen, die Lerner*innen, aber auch das Lernen selbst – etwa mit Blick auf die Didaktisierung – angepasst werden.

Während es selbstverständlich auch proprietäre Angebote speziell für die Bildung gibt, geht es hier v.a. darum, dass KI-Systeme durch ihre Nutzer*innen und an ihr konkretes Einsatzfeld angepasst werden können.

Ein praktisches Beispiel hierfür ist die Entwicklung spezialisierter KI-Labs für einen offenen Online-KI-Crashkurs der Universität zu Köln. Durch die Kombination eines eigens entwickelten ILIAS-Plugins sowie der Nutzung von Open Source-KI.nrw[5] als Inferenzdienst konnten an der Universität zu Köln nicht nur ein eigenes didaktisches Konzept exakt so umgesetzt werden wie gewünscht, sondern auch sichergestellt werden, dass Lernende und Lehrende der eingesetzten Technologie vertrauen und diese auf Wunsch auch vollständig nachvollziehen können[6].

Solche Entwicklungen zeigen, dass Hochschulen nicht auf »fertige« Lösungen angewiesen sind, sondern durch offene KI mehr Gestaltungsmacht, aber auch Vertrauen (vgl. Hannig et al. 2025), gewinnen können.

---

5 Open Source-KI.nrw ist ein Projekt der Universität zu Köln und der Ruhr-Universität Bochum zur Etablierung einer Open Source basierten KI-Infrastruktur.

6 Online-KI-Crashkurs: https://www.edulabs.uni-koeln.de/goto.php/crs/25328; ILIAS-Plugin: https://github.com/cce-uzk/AIChatPageComponent.

Offene KI und das Bildungs- und Forschungsverständnis

Ein zweites Argument betrifft die Passung offener KI mit einem modernen, offenen und zukunftsgerichteten Bildungs- und Forschungsverständnis. Hochschulen haben in den vergangenen Jahren immer stärker auf Offenheit gesetzt – etwa durch *Open Access*, *Open Educational Resources* oder *Open Science*. Offene KI fügt sich nahtlos in diese Entwicklung, welche die Verantwortung und den Auftrag öffentlicher Hochschulen betont, ein.

Sowohl die öffentliche Bildung als auch die Forschung zeichnen sich durch hohe Ansprüche an Souveränität, Transparenz und Nachvollziehbarkeit, Sicherheit und nicht zuletzt ethische und soziale Standards aus. Allein mit Blick auf die gute wissenschaftliche Praxis und das forschungsnahe Lehren und Lernen ist offene KI unabdingbar, da sie durch die ihr inhärenten Merkmale ebendiesen Ansprüchen genügen kann (vgl. Busse/Kleiber 2025: 183).

Kulturell betrachtet verkörpert offene KI sowohl das Ethos der Open-Source-Bewegung (Gupta 2024) als auch das der offenen Bildung und Wissenschaft. So verstanden ist offene KI nicht nur Technologie, sondern ein Gestaltungsprinzip für Bildung und Wissenschaft im 21. Jahrhundert.

Kompetenzaufbau durch offene KI

Drittens bietet offene KI Chancen für den Kompetenzaufbau. Während proprietäre Systeme häufig Blackboxes sind, erlaubt offene KI es, unter die Motorhaube zu schauen und zu experimentieren. Wir sehen didaktische *Affordances* (Kirschner et al. 2004) in ihrer Transparenz und Erklärbarkeit, Modifizier- und Anpassbarkeit, Integrier- und Erweiterbarkeit, Unmittelbarkeit und Isolierbarkeit sowie der Kollaboration, der Reflexion, dem Experimentieren, der Personalisierung und Individualisierung und der Chance auf inklusive und gerechte Didaktik, z.B. hinsichtlich des Zugangs.

Der Einsatz offener KI entfaltet dabei nicht nur Potenziale für technisch-informatorische oder KI-bezogene Kompetenzen – etwa durch den Bau oder die Anpassung offener KI-Systeme durch Lernende – sondern auch für andere fachbezogene Kompetenzen. So können etwa Risiken von KI, etwa mit Blick auf Verzerrungen oder Manipulation, an selbstbetriebenen und ungefilterten KI-Modellen, etwa im Vergleich mit proprietären Modellen, erleb-, versteh- und reflektierbar gemacht werden. Ebenso ist es denkbar, in Sprachwissenschaften die (kognitive) Modellierung von Sprache anhand von offenen Modellen zu thematisieren und Unterschiede und Gemeinsamkeiten menschlicher und künstlicher Intelligenz zu beleuchten.

Fazit und Ausblick

Offene KI ist mehr als eine minderwertige Alternative zu proprietärer KI und bietet Chancen für Bildungsprozesse. Gerade vor dem Hintergrund immer leistungsfähigerer offener KI plädieren wir dafür, sie nicht nur durch die Brille der Kosten, Leistungsfähigkeit oder Souveränität, sondern auch als Chance für Bildungs- und Transformationsprozesse zu verstehen. In der Diskussion zum Beitrag zeigte sich aber deutlich die Problematik der Kosten offener KI bei gleichzeitig – im Vergleich zu kommerziellen Angeboten – augenscheinlich geringeren Leistungsfähigkeit.

Wir verstehen offene KI ungeachtet dessen als Chance für Hochschulen, Mitgestalter*innen einer offenen und guten KI-Zukunft zu sein, welche zu den Werten und Prinzipien öffentlicher Bildung und Forschung passt. Zur Realisierung dieser Idee braucht es aber mindestens das explizite Commitment von Hochschulen zur Offenheit sowie enge Zusammenarbeit, wie sie in den NRW-weiten Projekten bereits gelebt wird.

Literatur

Busse, Beatrix/Kleiber, Ingo (2025): KI in der Hochschulbildung, München: Pearson Studium.

Eiras, Francisco/Petrov, Aleksandar/Vidgen, Bertie/Schroeder, Christian/Pizzati, Fabio/Elkins, Katherine/Mukhopadhyay, Supratik/Bibi, Adel/Purewal, Aaron/Botos, Csaba/Steibel, Fabro/Keshtkar, Fazel/Barez, Fazl/Smith, Genevieve/Guadagni, Gianluca/Chun, Jon/Cabot, Jordi/Imperial, Joseph/Nolazco, Juan A./Landay, Lori/Jackson, Matthew/Torr, Phillip H. S./Darrell, Trevor/Lee, Yong/Foerster, Jakob (2024): Risks and Opportunities of Open-Source Generative AI, https://arxiv.org/abs/2405.08597 vom 29.05.2024.

Gupta, Arun (2024): Fostering Open Source Culture. Increase Innovation and Deliver Faster with Open Source, Berkeley: Apress.

Hannig, Leon/Bush, Annika/Aksoy, Meltem/Trappen, Tim/Becker, Steffen/Ontrup, Greta (2025): Campus AI vs. Commercial AI: How Customizations Shape Trust and Usage of LLM as-a-Service Chatbots, http://arxiv.org/pdf/2509.15826v2 vom 23.09.2025.

Kirschner, Paul/Strijbos, Jan-Willem/Kreijns, Karel/Beers Pieter Jelle (2004): »Designing Electronic Collaborative Learning Environments«, in: Educational Technology Research and Development 52, S. 47–66.

Machado, Jorge (2025): Toward a Public and Secure Generative AI. A Comparative Analysis of Open and Closed LLMs, https://arxiv.org/abs/2505.10603

Paaßen, Benjamin (2025): Open Models statt OpenAI, https://www.jmwiarda.de/blog/2025/09/23/open-models-statt-openai vom 23.09.2025.

Vake, Domen/Šinik, Bogdan/Vičič, Jernej/Tošić, Aleksandar (2025): Is Open Source the Future of AI? A Data-Driven Approach, https://arxiv.org/abs/2501.16403 vom 27.01.2025.

Das Sprechen über KI und AI Literacy

Didaktische Konsequenzen aus einer Korpusanalyse[1]

Sophia Hercher[2]

Versteht generative KI, was sie da sagt? Eine intuitive Betrachtung des Sprachgebrauchs im Zusammenhang mit sprachgenerativer KI legt ein hohes Maß an Anthropomorphisierung nahe, das ist die Zuschreibung menschlicher Qualitäten auf nichtmenschliche Dinge. Diese spiegelt sich in unserer Nutzung der Systeme wider. Durch die Zuschreibung, dass generative KI die Bedeutung des Outputs kennt, schreiben wir beispielsweise den Texten einen Informations- oder sogar Wahrheitsgehalt zu. Dies führt jedoch zu einem falschen Verständnis der KI-Funktionsweise und zieht potenziell eine ungünstige Entwicklung der KI-Kompetenz nach sich.

In diesem Beitrag wird die linguistische Repräsentation des Verständnisses text-generativer KI mit Methoden der Korpuslinguistik untersucht, um die intuitive Betrachtung empirisch zu überprüfen. Hieraus werden Konsequenzen für die didaktische und sprachliche Gestaltung von Aufgaben in der Hochschullehre, bei denen text-generative KI zum Einsatz kommt, abgeleitet. Im Zentrum steht dabei die originäre Nutzungsweise und der konstruktive Umgang mit dem Output sowie die Betonung der aktiven Rolle Lernender.

Talking about AI and AI literacy – didactic design consequences from a corpus analysis

Does text-generative AI understand its own output? Looking at how we talk about generative AI suggests a high degree of anthropomorphism, that is the attribution of human characteristics onto non-human things. The way we speak about the AI systems can manifest in the way we use them. Assuming that generative AI knows the meaning of its output, we ascribe the generated texts informative value or substance. This, however, may lead to a wrong understanding of its functionality and a resulting adverse development of AI literacy.

This contribution studies the linguistic representation of our understanding of text-generative AI utilizing methods from corpus linguistics in order to verify intuitive assumptions about general language use empirically. From the results, consequences for higher education didactics and university teaching are drawn, specifgically for the design of tasks that include the

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0009–0008-6018-1161

usage of text-generative systems. At the center of the recommendations are the focus on the original design of the technology, the critical interaction with output as well as the emphasis of learners' active roles in the process.

Die Darstellung eines Chatbots als Mensch

Die Nähe zu menschlicher Sprache, die im Output textgenerativer KI zu erkennen ist, wirkt gleichermaßen eindrucksvoll wie potenziell besorgniserregend. In einer Studie zur linguistischen Repräsentation konzeptueller Metaphern gehe ich der Frage nach, ob ChatGPT generell anthropomorphisiert[3] wird und in welchen Kontexten dies geschieht (Hercher, 2026, angenommen). Im vorliegenden Beitrag umreiße ich die Kernannahmen und -ergebnisse dieser Arbeit, um Ableitungen für die Hochschuldidaktik und Hochschullehre vorzunehmen.

Ausgangspunkt für die Untersuchung war die Vermutung zahlreicher Autor*innen, dass die Anthropomorphisierung von KI in einem Missverständnis der Technologie (vgl. Bender/Koller 2020: 5186; Bender et al. 2021: 617) und entsprechend ungeeigneter Nutzungsweisen resultieren könnte (vgl. Hunger 2023: 4; Shanahan 2024: 79). Stellt man dies der Forderung nach einer reflektierten und konstruktiven Nutzungsweise von KI-Systemen in der Bildung gegenüber, ergibt sich die Frage, ob zur KI-Kompetenz auch eine bestimmte Weise des Sprechens über KI gehört, z.B. in einer nicht-anthropomorphisierenden Weise. Und wie könnte Kommunikation in der Bildung gestaltet sein, um diese möglichen negativen Konsequenzen zu vermeiden?

Die Forderung nach der Reduzierung von Anthropomorphismen (Hunger 2023) erinnert an die Metapherntheorie, welche ihren Ursprung in den 1980ern mit dem Werk Metaphors We Live By hat (vgl. Lakoff/Johnson 1980). Die Autoren zeigen, dass die Weise, in der wir sprechen, Aufschluss über unser Verstehen geben kann. Metaphern sind demnach nicht lediglich rhetorische Mittel, sondern essenzieller Teil unserer kognitiven Strukturen. So würde zum Beispiel der Satz »Solide Annahmen bilden das Fundament meiner Theorie.« einen Hinweis auf die konzeptuelle Metapher THEORIEN SIND GEBÄUDE[4] geben. Die Weise, in der wir Sprache nutzen, kann also Aufschluss darüber geben, wie wir Dinge verstehen (vgl. Lakoff/Johnson 1980: 5). Im Beispiel werden für Theorien ähnliche oder identische Konzepte benutzt wie für Gebäude, ausgedrückt durch die Worte Fundament und solide. Die Autoren erklären, dass die Übertragung von etwas uns Vertrautem (Menschsein) auf etwas weniger Vertrautes oder Abstraktes (Computer) stattfindet (vgl. Lakoff/Johnson 1980:

3 Ein Anthropomorphismus ist die Zuschreibung menschlicher Eigenschaften auf nichtmenschliche Dinge.

4 Konzeptuelle Metaphern werden konventionell in Kapitälchen geschrieben.

105). Bereits 1991 unternahmen Lakoff et al. den Versuch, eine umfassende Liste solcher konzeptuellen Metaphern aufzustellen. Hier findet sich auch die konzeptuelle Metapher COMPUTERS/PROGRAMMS ARE INTELLIGENT PEOPLE WITH INDEPENDENT WILLS. Dies deckt sich mit der Beobachtung, dass Anthropomorphisierung eine wichtige Rolle bei der Definition von neuen Technologien spielt (vgl. Carbonell/Sánchez-Esguevillas/Carro 2016: 148).

Anthropomorphismen von KI sind in vielen Bereichen prävalent: Systemseitig werden z.B. »Denkprozesse« in animierter Weise dargestellt (vgl. Abercrombie et al. 2023: 4778) oder der Bot »spricht« in der ersten Person singular. Skulmowski (2024) erwähnt in seiner Übersicht über die Effekte von KI-Anthropomorphisierung die Reduktion psychologischer Distanz zum System (vgl. Li/Sung 2021: 6) und die Steigerung der Kundenzufriedenheit (vgl. Janson 2023: 11). Demnach hebt Anthropomorphisierung tendenziell eher positive Eigenschaften eines Systems hervor.

Mit der Annahme, dass Sprache Aufschluss über die Denkstrukturen gibt, geht auch einher, dass linguistische Repräsentation sich auf das Verständnis eines Konzepts auswirken kann (vgl. Lakoff/Johnson 1980: 34). Hieran knüpfen die eingangs genannten Befürchtungen an: Durch übermäßige Anthropomorphisierung könnte ein falsches Verständnis der Funktionalität der Technologie entstehen. Dies reicht von unangemessenem Vertrauen in die »Fähigkeiten« von KI-gestützten Chatbots, über die daraus resultierende Nutzung der Systeme für Zwecke, für die sie nicht entwickelt wurden[5], bis hin zu einer möglichen Verantwortungsdiffusion durch die Attribution von Willenskraft oder Eigenständigkeit des Chatbots.

Um Konsequenzen für die Hochschullehre aus diesem Sprachgebrauch abzuleiten, ist zunächst zu identifizieren, wie tiefgreifend die Anthropomorphisierung von KI ist und in welchem Zusammenhang sie speziell auftaucht.

Methodischer Zugang und Datenbeispiele

In der Linguistik ist es üblich, Intuitionen über Sprachgebrauch mit großen Mengen an Sprachdaten zu überprüfen, so auch in der Metaphernforschung (z.B. Deignan 2005). Hierfür werden Sprachkorpora genutzt, also große Sammlungen an authentischem, repräsentativem Sprachmaterial. Für die hier referenzierte Untersuchung wurde der *News on the Web Corpus* (NOW, Davies 2016) genutzt, der umfassende und tagesaktuelle Sprachdaten in Form von Nachrichtentexten aus dem Internet enthält. Er ist für das Vorhaben gut geeignet, weil die enthaltenen Nachrichtentexte in der Regel eine große Reichweite haben und als wichtige Quelle für Informatio-

5 vgl. Bender und Kollers Ausführungen zur »Bedeutung« von Output textgenerativer KI, S. 5186f.

nen und die öffentliche Meinungsbildung angesehen werden (vgl. Krennmayr 2015: 531).

Für die Untersuchung wurde nach Konstrukten der Form »ChatGPT VERB« im Zeitraum November 2022 bis November 2024 gesucht, um Personifikationen ausfindig zu machen. Nach der Datenbereinigung wurden aus den 100 häufigsten Verbkonstruktionen 4052 Konkordanzen (Treffer inkl. Kurzkontext) extrahiert. Typische Beispiele aus dem Datenset sind:

1. »A lawyer is currently facing harsh consequences in court after ChatGPT **made up** six cases that the lawyer cited without first verifying case details …« (2023–06–09, US, arstechnica.com)
2. »… ChatGPT **knows** what fact checking means.« (2023–08–16, NG, staradvertiser.com)
3. »… ChatGPT **refused** to write a poem about Donald Trump, referring to the president as a [a model of hate speech]« (2023–02–15, US, nypost.com)
4. »… ChatGPT **recommended** Etsy, Amazon and Zales – again, nothing I couldn't have thought [of myself].« (2024–11–12, AU, cnet.com)

Anhand der Beispiele lässt sich die Vielseitigkeit der Anthropomorphismen in den Nachrichtentexten ablesen. In Beispiel 1 wird es so dargestellt, als ob sich das System etwas ausdenken könne. Nicht nur die Annahme, dass ChatGPT derart kreativ sei, auch die Information dazu, dass ein Anwalt solche Inhalte bereits aktiv in einer Gerichtsverhandlung genutzt hat, zeigen auf, welche Konsequenzen unangemessenes Vertrauen in die Richtigkeit des KI-generierten Outputs haben kann. In Beispiel 2 wird nicht nur behauptet, dass ChatGPT irgendetwas wisse, sondern gleichzeitig, dass der Wissensinhalt sich auf das Überprüfen von Fakten beziehe. Als System, welches Worte auf Basis zuvor berechneter Wahrscheinlichkeiten aneinanderreiht, ist es ChatGPT nicht möglich, Fakten zu überprüfen, da es kein Konzept von »Bedeutung« besitzt (vgl. Shanahan 2024: 72; Bender/Koller 2021). Die Beschreibung des Outputs als Weigerung in Beispiel 3 unterstellt dem System, dass es nicht nur Kenntnis der Bedeutung des generierten Outputs hat, sondern auch eigene Willensstärke, die es nutzen kann, um einer Aufforderung nicht nachzukommen. Hier wird die Programmierung des Systems mit dessen Intention gleichgesetzt. Das Beispiel verdeutlicht zudem den systeminhärenten Bias, der durch die Vorauswahl von Datensätzen entsteht. Man kann es gut oder schlecht finden, dass das System bestimmte Outputs nicht generiert und es gibt gewiss auch Texte, die nicht produziert werden sollten. Doch besonders das dritte Beispiel zeigt, welche Kontrolle über die Ausgabe ausgeübt werden kann und schon wird (siehe auch Morgan/Shepardson über den *regulation ban* der US-Regierung). Dass die Ausgaben von KI-Chatsystemen zum Beispiel in Form von Namensnennungen bestimmter Firmen kommerzialisiert werden

(vgl. Bsp. 4), stellt eine weitere Ausprägung der Informationskontrolle dar, in der Anthropomorphismen in Form persönlicher Empfehlungen genutzt werden.

Dies sind nur vier typische Beispiele, in denen einem KI-System Wissen über die Bedeutung von Text-Output, Verständnis des Inhalts, eigene Willenskraft oder Meinung zugeschrieben werden. Insgesamt konnten etwa die Hälfte der analysierten Konkordanzen als Personifikationen klassifiziert werden. Zu etwa einem Drittel waren dies sehr starke Zuschreibungen in der eben beschriebenen Form. Weiterhin bezogen sich die Personifikationen überwiegend auf den Output bzw. die Interaktion mit dem Chatbot, die i.d.R. das Verstehen des Inhalts seitens des Bots implizierten. Somit kann gefolgert werden, dass Anthropomorphismen prävalent sind und es insbesondere das Beschreiben der Funktionalität ist, für das sie genutzt werden.

Diskussion und Anwendung auf die Hochschullehre

Nun könnte man entgegenhalten, dass diese Beschreibungen eher harmlos sind. Wie auch Shanahan (2024) hierzu hervorhebt, verschwimmen bei textgenerativen KI-Systemen jedoch die Grenzen zwischen Sprachgebrauch und resultierender Nutzung, weil die Sprache in beiden Fällen das Medium ist. Es ist wichtig und teilweise schwierig, zu verstehen, dass es nicht ChatGPT ist, das Empfehlungen für das Shopping ausspricht (Beispiel 4) oder sich weigert, ein Gedicht über Donald Trump zu schreiben (Beispiel 3), sondern die Personen und Firmen, welche die Entwicklung des Chatbots verantworten. Die Annahme eines neutralen und wohlwollenden Gesprächspartners ist nicht nur falsch, sondern potenziell gefährlich, weil Menschen zu Entscheidungen auf der Grundlage falscher Annahmen bewogen werden können.

Doch auch die Deanthropomorphisierung ist keineswegs trivial, denn oft kennen wir keine nicht-anthropomorphisierenden Begriffe in diesem Kontext. Sachangemessene Beschreibungen können ferner als umständlich empfunden werden (Shanahan/McDonell/Reynolds, 2023: 493). Zusätzlich ist zu beobachten, dass Prompt-Kurse oder Hilfeseiten nahelegen, emotionale Sprache zu nutzen (Mok, 2023) oder Rollen zuzuweisen, um bessere Outputs zu generieren (Schulhoff, 2024). Die menschenähnliche Kommunikationsweise erleichtert die Interaktion für die Nutzenden. Auch wenn die Ratschläge zur Nutzung sich fast noch schneller ändern als neue Modelle veröffentlicht werden, liegt hierin der zentrale Konflikt. Die Verbesserung des Outputs durch Anthropomorphisierung seitens der Nutzenden steht im direkten Gegensatz zur akkuraten Beschreibung und daraus resultierendem Verständnis des Systems.

Es wird deutlich, dass KI-Kompetenz auch in Bezug auf das Sprechen über KI mehrere Facetten beinhaltet. Lehrende sollten deshalb ein gutes Gefühl für die Angemessenheit bestimmter Ausdrucksweisen und die Etablierung akkurater Sprach-

nutzung entwickeln, um den Aufbau von KI-Kompetenz in diesem Bereich gut unterstützen zu können. Hierfür werden im Folgenden ein paar Vorschläge gemacht.

Ein wichtiger Bereich der Kommunikation, in dem Lehrende die Funktionsweise von KI-Systemen adressieren, ist die Aufgabenstellung. Wenn wir vermeiden wollen, dass aus der Anthropomorphisierung der Systeme eine unsachgemäße Nutzung resultiert, können wir in diesem Bereich besondere Sorgfalt aufwenden. Im Folgenden werden drei einfache Prinzipien des Aufgabendesigns vorgeschlagen, die bei der Entwicklung von KI-Kompetenz helfen sollen. Diese sind:

1. Betonung der aktiven Rolle der Lernenden in der Interaktion mit dem Chatbot
2. Die Hervorhebung des eigentlichen Zwecks des Systems (Texterstellung)
3. Die Förderung der kritisch-konstruktiven Interaktion mit Output

Um die aktive Rolle der Lernenden hervorzuheben, reicht es zum Beispiel schon, statt »Fragen Sie die KI nach Vorschlägen für eine Werbekampagne«, zu sagen »Ermitteln Sie unter Nutzung eines textgenerativen KI-Systems, Vorschläge für die Werbekampagne«. Hier wird der Werkzeug-Charakter hervorgehoben und die Nutzenden bleiben die Akteure.

Um die eigentliche Funktion zu nutzen in den Vordergrund zu stellen, wäre etwa denkbar, das System zur Herstellung von Textformen zu nutzen. Ein Beispiel hierfür könnte sein: »Erstellen Sie mit generativer KI ein Gedicht. Erörtern Sie, welche Weltansichten in der Ausführung erkennbar werden. Welche Rückschlüsse lässt das Ergebnis auf Trainingsdaten zu?« Eine solche Aufgabe hebt den textgenerativen Charakter des Systems hervor und leitet zur Reflexion der Trainingsdatensätze und deren Auswirkung auf die Textgenese ein. Besonders geeignet für die kritisch-konstruktive Interaktion mit dem Output sind auch Vergleichsaufgaben, in denen der gleiche Prompt in unterschiedliche KI-Systeme eingegeben wird, um die Qualität der Outputs begründet zu bewerten. Grundsätzlich sollten Lehrende Aufgaben vermeiden, bei denen eine Kenntnis oder ein Verständnis des Inhalts seitens des KI-Chatbots vorausgesetzt ist, z. B. Rechercheaufgaben, denn die Gleichsetzung generativer KI-Chatbots mit Suchmaschinen oder gar Datenbanken scheint eine sehr häufige zu sein, obgleich sie irreführend ist.

Fazit

Zusammenfassend lässt sich sagen, dass Anthropomorphismen ein fester Bestandteil des Sprechens über KI sind. Jedoch ist eine vollständige Deanthropomorphisierung des Sprechens über generative KI-Systeme wahrscheinlich sehr schwierig und vielleicht nicht einmal wünschenswert oder möglich, wenn es um die Gestaltung von Prompts geht. Die Prägung unserer Wahrnehmung und unseres Verstehens der

Systeme durch Anthropomorphismen hin zu unangemessenem Vertrauen in deren Outputs steht im Kontrast zur Steigerung der Qualität durch die Nutzung menschenähnlicher Sprache bei der Interaktion mit den Systemen. Jedoch haben wir als Lehrende die Möglichkeit, unsere Kommunikation so zu gestalten, dass die aktive Rolle Lernender in den Fokus gerückt wird und wir uns durch die bewusste Verwendung unserer Sprache regelmäßig die tatsächliche Funktionsweise der KI-Systeme in unser und das Bewusstsein der Lernenden rufen.

Literatur

Abercrombie, Gavin/Cercas Curry, Amanda/Dinkar, Tanvi/Rieser, Verena/Talat, Zeerak (2023): »Mirages. On Anthropomorphism in Dialogue Systems« in: Houda Bouamor/Juan Pino/Kalika Bali (Hg.), Proceedings of the 2023 Conference on Empirical Methods. in: Natural Language Processing Singapore: Association for Computational Linguistics, S. 4776–90.

Bender, Emily M./Gebru, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret (2021): »On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?«, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT ›21. New York, NY, USA: Association for Computing Machinery, S. 610–23. https://doi.org/10.1145/3442188.3445922.

Bender, Emily M./Koller, Alexander (2020): »Climbing Towards NLU: On Meaning, Form, and Understanding in the Age of Data« in: Dan Jurafsky/Joyce Chai/Natalie Schluter/Joel Tetreault (Hg.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, USA: Association for Computational Linguistics, S. 5185–98. https://doi.org/10.18653/v1/2020.acl-main.463.

Carbonell, Javier/Sánchez-Esguevillas, Antonio/Carro, Belén (2016): »The Role of Metaphors in the Development of Technologies. The Case of the Artificial Intelligence.«, in: Futures 84, S. 145–53. https://doi.org/10.1016/j.futures.2016.03.019.

Davies, Mark (2016): »Corpus of News on the Web (NOW).« https://english-corpora.org/now

Deignan, Alice (2005): Metaphor and Corpus Linguistics, Converging Evidence in Language and Communication Research. Amsterdam [u.a.]: Benjamins.

Hercher, Sophia (2026, angenommen): »What's ChatGPT Doing? – Towards a Metaphor Profile of Generative AI (Working Title).« Unveröffentlichter Beitrag, in Planung, in: Callies, Marcus und Peters, Arne. Cognitive Sociolinguistic Studies on Conceptualisation: Contexts, Data, and Methods (Working Title). Amsterdam.

Hunger, Francis (2023): »Unhype Artificial ›Intelligence‹! A Proposal to Replace the Deceiving Terminology of AI«, in: Zenodo, 1.

Janson, Andreas (2023): »How to Leverage Anthropomorphism for Chatbot Service Interfaces: The Interplay of Communication Style and Personification«, in: Computers in Human Behavior 149, S. 107954.

Krennmayr, Tina (2015): »What Corpus Linguistics Can Tell Us about Metaphor Use in Newspaper Texts«, in: Journalism Studies 16, 4, S. 530–46.

Lakoff, George/Espenson, Jane/Goldberg, Adele (1991): »Master Metaphor List.« http://araw.mede.uic.edu/alansz/metaphor/METAPHORLIST.pdf

Lakoff, George/Johnson, Mark (1980): Metaphors We Live by, Chicago, Ill. [u.a.]: Univ. of Chicago Press.

Li, Xinge/Sung, Yongjun (2021): »Anthropomorphism Brings Us Closer: The Mediating Role of Psychological Distance in User–AI Assistant Interactions«, in: Computers in Human Behavior, 118, S. 106680.

Morgan, David/Shepardson, David (2025): »US Senate Strikes AI Regulation Ban from Trump Megabill.«, in: https://www.reuters.com/legal/government/us-senate-strikes-ai-regulation-ban-trump-megabill-2025-07-01/ vom 01.06.2025.

Schulhoff, Sander (2024): »Rollen Prompting.«, in: https://learnprompting.org/de/docs/basics/roles

Shanahan, Murray (2024): »Talking about Large Language Models«, in: Commun. ACM, 67 (2), S. 68–79.

Shanahan, Murray/McDonell, Kyle/Reynolds, Laria (2023): »Role Play with Large Language Models.«, in: Nature 623, S. 493–98.

Skulmowski, Alexander (2024): »Placebo or Assistant? Generative AI Between Externalization and Anthropomorphization.«, in: Educational Psychology Review 36: 2, S. 58.

Förderung von AI Literacy bei Studierenden

Ein Modell für die Integration von KI-Kompetenzen in die Hochschulbildung[1]

Miriam Stehling[2], Claudia Berg[3], Manuel Geisler[4], Miriam Hägerbäumer[5], Nils Aschhoff[6] und Vera Lenz-Kesekamp[7]

Die fortschreitende Integration von Künstlicher Intelligenz (KI) in Bildung und Arbeitswelt erfordert eine Anpassung der Hochschulcurricula, um Studierende auf eine KI-geprägte Zukunft vorzubereiten. Dieser Beitrag beschreibt die Entwicklung eines entsprechenden KI-Kompetenzmodells für den Einsatz im Hochschulkontext und dessen anschließende Verankerung und Integration in die Lehrpraxis. Der Beitrag hebt durch die systematische Integration verschiedener Studien einerseits entscheidende Aspekte für die Förderung von AI Literacy im Hochschulkontext hervor und verdeutlicht, dass eine »One-size-fits-all«-Lösung im Kontext der KI-Kompetenzentwicklung nicht den Bedarfen einer zukunftsgerichteten Hochschulbildung entspricht. Andererseits liefert der Beitrag (lehr-)praxisorientierte Anknüpfungspunkte für die Entwicklung von Kompetenzmodellen und deren anschließende hochschulweite Integration.

Promoting AI Literacy among Students: A Model for Integrating AI Competencies into Higher Education

The advancing integration of Artificial Intelligence (AI) into education and the workplace necessitates an adaptation of university curricula to prepare students for an AI-driven future. This paper outlines the development of an AI competency model designed for application in the higher education context, and its subsequent embedding and integration into teaching practices. By systematically integrating various studies, the paper highlights crucial aspects for promoting AI literacy in higher education while illustrating that a »one-size-fits-all« solution

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0009–0003-8516-4484
3 ORCID: 0009–0008-0544-5646
4 ORCID: 0009–0003-8320-3240
5 ORCID: 0000–0003-4248-1570
6 ORCID: 0000–0003-4963-0751
7 ORCID: 0009–0004-9845-7953

does not meet the needs of future-oriented higher education in the context of AI competency development. Additionally, the paper provides practice-oriented hints for the development of competency models and their subsequent institution-wide integration.

Einleitung

Die fortschreitende Integration von Künstlicher Intelligenz (KI) in sämtliche Lebens- und Arbeitsbereiche stellt eine der tiefgreifendsten Transformationen des 21. Jahrhunderts dar (OECD 2023: 4). Insbesondere die breite Verfügbarkeit generativer KI-Systeme hat eine disruptive Dynamik entfaltet, die Bildungseinrichtungen, Arbeitsumgebungen und die Gesellschaft als Ganzes rasant verändert (Bobula 2024: 2–3; Cardon et al. 2023: 257–258). Diese Entwicklung konfrontiert Hochschulen mit der dringenden Notwendigkeit, ihre Curricula anzupassen und Studierende auf eine von KI geprägte Zukunft vorzubereiten (Laupichler et al. 2022: 14; Southworth et al. 2023: 3). Die Herausforderung besteht darin, über eine rein reaktive Auseinandersetzung hinaus proaktiv die Kompetenzen für einen souveränen, kritischen und produktiven Umgang mit KI zu definieren und zu vermitteln. Diese Kompetenzen werden unter dem Begriff der KI-Kompetenzen (*AI Literacy*) zusammengefasst. Sie gliedern sich ein in den breiteren Rahmen der oft unter dem Sammelbegriff »Future Skills« diskutierten Fähigkeitsprofile (vgl. Ehlers 2023; Suessenbach et al. 2021), die Kompetenzen betonen, die Handlungsfähigkeit auch in neuartigen Kontexten sicherstellen, wie etwa Problemlösungsfähigkeit, Adaptionsfähigkeit, kritisches Denken und digitale Kollaboration (Drewes/Hägerbäumer 2025: 4, 9–10). Wie die Ausbildung von KI-Kompetenzen auf Basis eines systematischen Kompetenzmodells im hochschulischen Kontext gelingen kann, wird nachfolgend thematisiert.

Grundlagen von KI-Kompetenzmodellen

KI-Kompetenzen ermöglichen es Individuen, die Funktionsweisen, Potenziale und Risiken von KI-Systemen zu verstehen, kritisch zu bewerten und produktiv mit ihnen zu interagieren (Long/Magerko 2020: 2). Die systematische Verankerung solcher Kompetenzen in Hochschulcurricula ist daher eine zwingende Notwendigkeit, um Absolventinnen und Absolventen zukunftsfähig auszubilden (Lambrecht et al. 2025: 33).

KI-Kompetenzen werden dabei als ein multidimensionales Konstrukt verstanden, das technisches Verständnis mit kritischer Reflexion und praktischer Anwendung verbindet (Cox 2024: 94; Ng et al. 2021: 1, 4). Long und Magerko (2020: 2) beschreiben sie als ein Set von Kompetenzen, das es Individuen ermöglicht, KI kritisch zu bewerten, effektiv mit ihr zu kommunizieren und sie als Werkzeug zu nutzen.

Dieser ganzheitliche Ansatz wird in der neueren Forschung bestätigt und erweitert. So betonen Allen und Kendeou (2024: 4), dass KI-Kompetenzen die Fähigkeit umfassen, mit KI auf produktive Weise zu kommunizieren und zu kollaborieren.

Die Integration von KI-Kompetenzen in die Hochschullehre ist entscheidend, um Studierende auf die Anforderungen des zukünftigen Arbeitsmarktes vorzubereiten (Bobula 2024: 15; Rampelt et al. 2025: 12). Während in der Forschung bereits erste Modelle und Skalen zur Erfassung von KI-Kompetenzen entwickelt wurden (z.B. Carolus et al. 2023; Hornberger et al. 2023; Laupichler et al. 2023), besteht die Notwendigkeit, diese Ansätze für spezifische Zielgruppen und Kontexte zu adaptieren. Ein spezifisches Modell muss den Anwendenden Spielraum für eigene Festlegungen darüber bieten, welche Kompetenzen in welcher Tiefe vermittelt werden sollen.

Eine wesentliche Grundlage für die inhaltliche Strukturierung von KI-Kompetenzen liefern die Arbeiten von Laupichler et al. (2022, 2023). In ihrer Scoping-Literaturübersicht arbeiteten Laupichler et al. (2022) den Forschungsstand zu AI Literacy in der Hochschul- und Erwachsenenbildung auf. Darauf aufbauend entwickelten Laupichler et al. (2023) eine validierte Skala zur Erfassung von KI-Kompetenzen bei Nicht-Expert:innen. Die empirischen Ergebnisse von Laupichler et al. (2023: 6) zeigen, dass sich KI-Kompetenzen in drei zentrale, voneinander abgrenzbare Kompetenzbereiche gliedern lassen: Technisches Verständnis, Kritische Bewertung und Praktische Anwendung.

Neben der inhaltlichen Gliederung nach Kompetenzbereichen ist eine Differenzierung nach Kompetenztiefe ebenso von Bedeutung. Ein »One-size-fits-all«-Ansatz würde der Heterogenität der Studierenden und den unterschiedlichen Anforderungen verschiedener Fachdisziplinen und Berufsfelder nicht gerecht. Hierfür bietet der Ansatz von Teuber, Dindarian und Ekaterina (2022) einen entscheidenden konzeptionellen Anker. Die Autor:innen argumentieren, dass die Anforderungen an den KI-Kompetenzerwerb je nach Rolle und Interaktionstiefe mit der Technologie variieren. Sie schlagen eine Differenzierung in drei Ebenen vor: grundlegende KI-Basiskompetenzen für die breite Bevölkerung, vertiefte Kompetenzen für die KI-Nutzung im beruflichen Kontext und spezialisierte Kompetenzen für KI-Expert:innen. Diese Unterscheidung ermöglicht es, Lernziele und -inhalte bedarfsgerecht zu gestalten und sowohl eine breite Grundbildung als auch fachspezifische Vertiefungen zu ermöglichen (Teuber et al. 2022: 100–101, 104–105).

Entwicklung eines bedarfsgerechten KI-Kompetenzmodells

In der hochschulischen Praxis stellt sich die Frage, wie bedarfsgerechte KI-Kompetenzen bei einer heterogenen Studierendenschaft systematisch und in der Breite gefördert werden können. Um diese Herausforderung zu adressieren, wurde an

der Europäischen Fernhochschule Hamburg (Euro-FH) in einem interdisziplinären Projektteam, bestehend sowohl aus Vertreter:innen aus dem Bereich der Lehre als auch aus dem Studienbetrieb, ein spezifisches KI-Kompetenzmodell entwickelt. Der Entwicklungsprozess basierte auf einer fundierten Literaturrecherche zu bestehenden Frameworks zu KI-Kompetenzen sowie auf iterativen Verfeinerungszyklen. In diesen wurden einzelne (Teil-)Kompetenzen des ausgearbeiteten Modells unter Berücksichtigung von Expert:innenbewertungen und dem Feedback von Stakeholder:innen (z.B. Lehrende) ausdifferenziert und definiert.

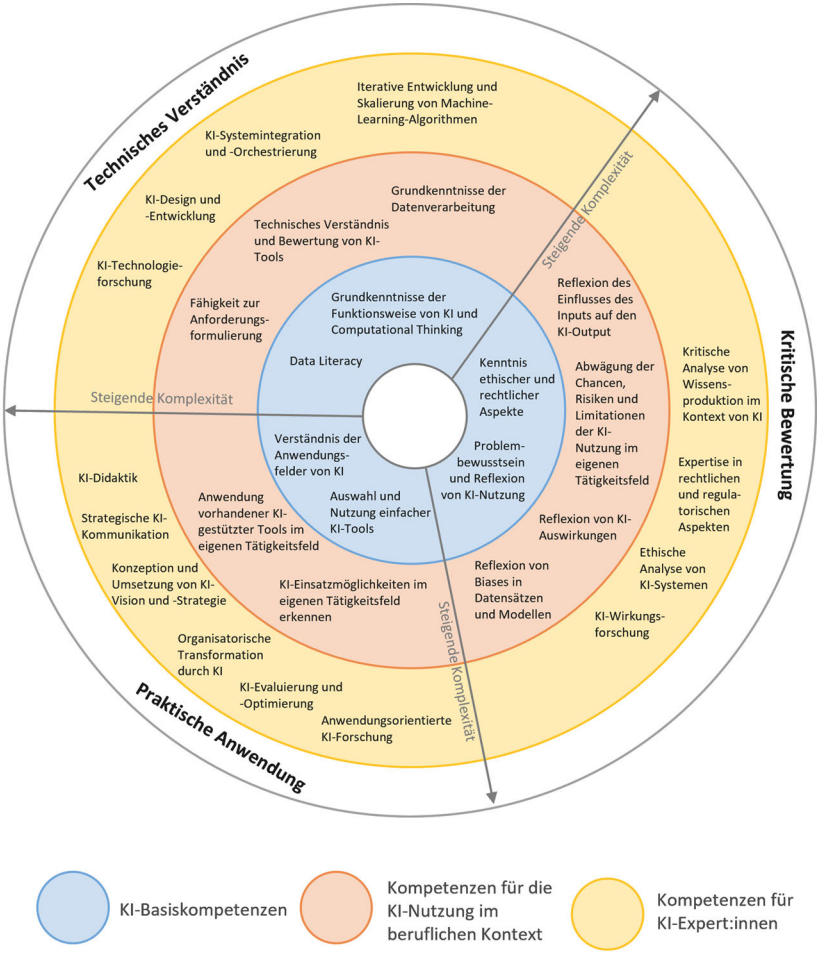
Das entwickelte Modell greift etablierte Ansätze auf und synthetisiert diese zu einem neuen Mehrebenenmodell. Die Gliederung in die **drei Sektoren** *Technisches Verständnis, Kritische Bewertung* und *Praktische Anwendung* lehnt sich an die faktorenanalytisch bestätigten Kompetenzbereiche von Laupichler et al. (2023: 6) an, wobei ethische Fragestellungen integraler Bestandteil der kritischen Reflexion von Handlungen, Entscheidungen und ihren gesellschaftlichen Auswirkungen sind. Diese Struktur gewährleistet, dass sowohl die technologischen Grundlagen als auch die kritisch-reflexive und die anwendungsorientierte Dimension von KI-Kompetenzen gleichwertig Berücksichtigung finden. Für die vertikale Differenzierung des Modells wurde der Ansatz von Teuber et al. (2022: 100) adaptiert, welcher Kompetenzanforderungen für unterschiedliche Zielgruppen unterscheidet. Unser Modell überträgt diesen Gedanken auf den Hochschulkontext und definiert **drei aufeinander aufbauende Kompetenzniveaus**: *die KI-Basiskompetenzen, den professionellen KI-Einsatz und das KI-Expert:innenlevel.* Diese Differenzierung ermöglicht es, KI-Kompetenzen bedarfsgerecht auszubilden – von einer grundlegenden AI Literacy für alle Studierenden bis hin zu spezialisierten Kompetenzen für bestimmte Anwendungsfelder.

Das entwickelte Modell betont die sektorübergreifende Verflechtung von Kompetenzen, welche sich gegenseitig bedingen und ergänzen. Die einzelnen Dimensionen des Modells werden im Folgenden genauer erläutert. Drei Sektoren bilden die inhaltlichen Säulen des Modells (siehe Abb. 1).

Der Sektor »**Technisches Verständnis«** bildet die Wissensgrundlage für einen effektiven und informierten Umgang mit KI. Ohne dieses Fundament bleibt die Nutzung von KI oberflächlich, eine kritische Bewertung ist kaum möglich. Der Sektor »**Kritische Bewertung«** ist für einen verantwortungsvollen KI-Einsatz unverzichtbar. Er bündelt Kompetenzen, die zur Reflexion über die Implikationen von KI befähigen. Im Sektor »**Praktische Anwendung«** wird die Fähigkeit geschult, KI als produktives Werkzeug im akademischen und beruflichen Kontext einzusetzen.

Drei differenzierbare Niveaus ermöglichen darüber hinaus eine passgenaue curriculare Integration und tragen der Heterogenität der Studierenden sowie der unterschiedlichen Relevanz von KI in verschiedenen Fachdisziplinen Rechnung (siehe ebenfalls Abb. 1).

Abbildung 1: KI-Kompetenzmodell der Euro-FH



1. **Grundlegende KI-Kompetenzen**: Dieses Niveau bildet die Basis und richtet sich an alle Studierenden, unabhängig von ihrem Fachgebiet. Es umfasst ein grundlegendes Verständnis der Funktionsweise von KI, essenzielle Data Literacy und ein Bewusstsein für zentrale Chancen und Risiken.
2. **Kompetenzen für den professionellen Einsatz von KI**: Dieses Niveau zielt auf Studierende, die KI-Werkzeuge in ihrem zukünftigen Berufsfeld aktiv nutzen werden. Die Kompetenzen sind hier stärker auf den spezifischen beruflichen Kontext zugeschnitten und umfassen die fortgeschrittene Anwendung von KI-Tools, ein erweitertes technisches Verständnis sowie eine vertiefte ethische Reflexion im jeweiligen Fachbereich.

3. **Kompetenzen für KI-Expert:innen**: Das höchste Niveau richtet sich an Studierende in spezialisierten, KI-nahen Studiengängen oder Vertiefungen. Es beinhaltet fortgeschrittene technische Fähigkeiten und das Spezialwissen, das für die Konzeption, Implementierung und das Management von KI-Systemen erforderlich ist.

Die Verknüpfung von Sektoren und Niveaus führt zu insgesamt 29 verschiedenen Kompetenzen mit entsprechend präzise ausformulierten Kompetenzbeschreibungen, die als Grundlage für die Formulierung von Lernzielen dienen.

Verankerung in der Hochschulpraxis

Die Herausbildung der KI-Kompetenzen erfordert eine entsprechende Integration des Modells in die Studiengangs- und Modulentwicklung der Hochschule. Neben der fortlaufenden Entwicklung des Kompetenzmodells befasste sich das Projektteam daher auch mit der Umsetzung des Modells in die Lehrpraxis an der EURO-FH. Im Ergebnis wurde eine interaktive Plattform geschaffen, welche den Lehrenden einen strukturierten Zugang zum Kompetenzmodell, den dazugehörigen Kompetenzdefinitionen sowie zu bereits im praktischen Einsatz erprobten Lehrmaterialien ermöglicht. Diese Plattform unterstützt Lehrende darin, die Anforderungen an KI-Kompetenzen zu verstehen und systematisch geeignete Lehrmaterialien für in Modulen und Studiengängen relevante KI-Kompetenzen auszuwählen.

Im Sinne eines Best Practice-Ansatzes können Lehrende außerdem eigene, bereits in der Lehrpraxis erprobte Lehrmaterialien zum Thema KI auf der Plattform teilen. Mittels eines intuitiven Meldeformulars übermitteln die Lehrenden Informationen zu den Lehrmaterialien und können diese einzelnen Kompetenzen des Modells zuordnen. Kuratiert vom Projektteam werden die Lehrmaterialien anderen Lehrenden dann auf der Plattform vorgestellt. Die Spannweite reicht von Prüfungsaufgaben über verschriftlichte Lehrmaterialien sowie Seminarkonzeptionen bis hin zu kompletten Modulen. Die Plattform fördert somit auch den Austausch zwischen den Lehrenden (Cardon et al. 2023: 281).

Im Projekt wurden zudem unter Berücksichtigung unterschiedlicher Taxonomiestufen Beispielformulierungen zur Integration der KI-Kompetenzen in die Modulbeschreibungen entwickelt. Den modulverantwortlichen Lehrenden bieten diese Beispielformulierungen Orientierung bei der Entwicklung von modulspezifisch passgenauen Lernzielen.

Diskussion

Der Diskurs zur AI Literacy im Hochschulkontext verdeutlicht die Notwendigkeit eines strategischen und integrativen Ansatzes zur Ausbildung von Kompetenzen in der Hochschulbildung. Der vorliegende Beitrag knüpft an diesen Diskurs an und zeigt mit der Entwicklung und Implementierung des Kompetenzmodells der Euro-FH einen Ansatz für die Systematisierung und Integration von KI-Kompetenzen im Hochschulkontext. Das Modell synthetisiert dabei die beiden Perspektiven einer inhaltlichen Breite nach Laupichler et al. (2023) und einer anwendungsorientierten Tiefe nach Teuber et al. (2022) und adressiert somit das Problem, dass eine »One-size-fits-all«-Lösung im Kontext der KI-Kompetenzentwicklung nicht der Verantwortung einer an den Bedarfen der Studierenden ausgerichteten Hochschulbildung entspricht. Insbesondere im Kontext einer Fernhochschule mit einem breiten fachlichen Portfolio und einer ausgeprägt heterogenen Studierendenschaft ist dieser strukturierte und zugleich differenzierte Ansatz essenziell.

Hervorzuheben ist allerdings, dass sich das entwickelte KI-Kompetenzmodell explizit auf den Einsatz im Hochschulkontext bezieht. Die Übertragbarkeit des Modells auf weitere Bildungskontexte ist kontextbezogen zu prüfen. Zudem wurde das Modell aus einer hohen berufspraktischen Orientierung heraus und mit einem Fokus auf die Anwendbarkeit in der Lehre entwickelt. Eine umfassende empirische Validierung steht aus.

Zukünftig sollen daher die Weiterentwicklung und die Bewertung von Anwendbarkeit, Passgenauigkeit und Wirksamkeit des Modells sowie der dazugehörigen interaktiven Plattform weiter beforscht und institutionalisiert werden. Die dynamische Anpassung des Modells und die folgende Integration in die Curricula der Hochschule sollen sicherstellen, dass Absolventinnen und Absolventen nicht nur als passive Anwender:innen, sondern als kompetente, kritische und verantwortungsbewusste Gestalter:innen der KI-getriebenen Zukunft agieren können.

Literatur

Allen, Laura Kristen/Kendeou, Panayiota (2024): »ED-AI Lit: An Interdisciplinary Framework for AI Literacy in Education«, in: Policy Insights from the Behavioral and Brain Sciences 11(1), S. 3–10.

Bobula, Michal (2024): »Generative artificial intelligence (AI) in higher education: a comprehensive review of challenges, opportunities, and implications«, in: Journal of Learning Development in Higher Education 30, S. 1–27.

Cardon, Peter/Fleischmann, Carolin/Aritz, Jolanta/Logemann, Minna/Heidewald, Jeanette (2023): »The Challenges and Opportunities of AI-Assisted Writing: De-

veloping AI Literacy for the AI Age«, in: Business and Professional Communication Quarterly 86(3), S. 257–295.

Carolus, Astrid/Koch, Martin J./Straka, Samantha/Latoschik, Marc Erich/Wienrich, Carolin (2023): »MAILS – Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies«, in: Computers in Human Behavior: Artificial Humans 1(100014), S. 1–10.

Cox, Andrew (2024): »Algorithmic Literacy, AI Literacy and Responsible Generative AI Literacy«, in: Journal of Web Librarianship 18(3), S. 93–110.

Drewes, Sylvana/Hägerbäumer, Miriam (2025): »Future Skills in der Arbeitswelt 4.0: Perspektiven zur Relevanz und Entwicklung von Future Skills im Arbeitskontext«, in: Miriam Hägerbäumer/Udo Thelen/André Renz (Hg.), Future Skills in Human Resource Management und Corporate Learning, Wiesbaden: Springer Gabler, S. 3–30.

Ehlers, Ulf-Daniel (2023): »Future Skills Framework. Kompetenzen für die KI-Welt«, in: managerSeminare 305, S. 70–77.

Hornberger, Marie/Bewersdorff, Arne/Nerdel, Claudia (2023): »What do university students know about Artificial Intelligence? Development and validation of an AI literacy test«, in: Computers and Education: Artificial Intelligence 5(100165), S. 1–12.

Lamprecht, Gabriella-Maria/Lintfert, Britta/Martschiske, Regine/Wiehenbrauk, Daniela (2025): »KI meets Lehre – Die Notwendigkeit der curricularen Verankerung von KI-Kompetenzen«, in: Zeitschrift für Hochschulentwicklung 20(SH-KI-2), S. 13–37.

Laupichler, Matthias Carl/Aster, Alexandra/Haverkamp, Nicolas/Raupach, Tobias (2023): »Development of the »Scale for the assessment of non-experts' AI literacy« – An exploratory factor analysis«, in: Computers in Human Behavior Reports 12(100338), S. 1–10.

Laupichler, Matthias Carl/Aster, Alexandra/Schirch, Jana/Raupach, Tobias (2022): »Artificial intelligence literacy in higher and adult education: A scoping literature review«, in: Computers and Education: Artificial Intelligence 3(100101), S. 1–15.

Long, Duri/Magerko, Brian (2020, 25.-30. April): »What is AI Literacy? Competencies and Design Considerations«, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems Honolulu HI USA, S. 116.

Ng, Davy Tsz Kit/Leung, Jac Ka Lok/Chu, Samuel Kai Wah/Qiao, Maggie Shen (2021): »Conceptualizing AI literacy: An exploratory review«, in: Computers and Education: Artificial Intelligence 2(100041), S. 1–11.

OECD (2023): OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market, Paris: OECD Publishing.

Rampelt, Florian/Klier, Julia/Kirchherr, Julian/Ruppert, Raffael (2025): KI-Kompetenzen in deutschen Unternehmen. Schlüssel zu einer Jahrhundertchance für Deutschland, Essen: Stifterverband für die Deutsche Wissenschaft e. V.

Southworth, Jane/Migliacchio, Kati/Glover, Joe/Glover, Ja'Net/Reed, David/McCarty, Christopher/Brendemuhl, Joel/Thomas, Aaron (2023): »Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy«, in: Computers and Education: Artificial Intelligence 4(100127), S. 1–10.

Suessenbach, Felix/Winde, Mathias/Klier, Julia/Kirchherr, Julian (2021): Future Skills 2021. 21 Kompetenzen für eine Welt im Wandel, Essen: Stifterverband für die Deutsche Wissenschaft e. V.

Teuber, Katharina/Dindarian, Azadeh/Ekaterina, Naz Cilo-van Norel (2022): »Künstliche Intelligenz und ihre Anforderungen an den Kompetenzerwerb«, in: Ralf Knackstedt/Jürgen Sander/Jennifer Kolomitchouk (Hg.), Kompetenzmodelle für den Digitalen Wandel. Orientierungshilfen und Anwendungsbeispiele, Berlin, Heidelberg: Springer, S. 99–115.

Zhang, Ceciley Xinyi/Rice, Ronald E./Wang, Laurent H. (2024): »College students' literacy, ChatGPT activities, educational outcomes, and trust from a digital divide perspective«, in: New Media & Society, 00(0), S. 1–23.

Lehr-/Lernangebote zum KI-gestützten Erwerb Akademischer Schreibkompetenz in Fremdsprachen[1]

Torsten Kirchner und Géraldine Stich-Desmarchelier

Die zunehmende Verbreitung von KI-Tools verändert akademische Schreibprozesse grundlegend und wirft die Frage auf, wie KI-Tools effektiv in die Fremdsprachenlehre integriert werden können. Das Sprachlernzentrum der Universität Bonn hat daraufhin ein Projekt zur KI-gestützten Förderung der akademischen Schreib- und Lesekompetenz Studierender in den Fremdsprachen Englisch, Französisch, Spanisch sowie Deutsch für die Niveaustufen B2+ und C1 (GER) entwickelt.
Die entwickelten Module sind im bereits bestehenden Programm Begleitetes Autonomes Fremdsprachenlernen verortet. In individuell zu bearbeitenden E-Learning-Aufgaben werden Chatbots etwa als persönlicher Feedbackgeber eingesetzt. Die Dokumentation des Lernprozesses sowie die kritische Reflexion über KI-Nutzung und -Output erfolgen in einem E-Portfolio. Erste Hinweise aus der Praxis deuten auf Potenziale hin und zeigen Herausforderungen, ohne Anspruch auf Repräsentativität zu erheben.

Teaching and learning resources for the AI-assisted acquisition of academic writing skills in foreign languages
The increasing prevalence of AI tools is changing academic writing processes fundamentally. This raises the question how such tools can be effectively and judiciously integrated into foreign language teaching. In response, the Language Centre at the University of Bonn has implemented a project specifically aiming to promote academic reading and writing skills of students in the foreign languages English, French, Spanish and German, for the levels B2+ and C1 (CEFR).
These developed modules are part of the existing Accompanied Autonomous Foreign Language Learning programme. While completing e-learning assignments individually, students use chatbots, for example, as personal feedback providers. They also keep an e-portfolio to document and reflect on their individual AI-enhanced learning process as well as the generated AI output. Preliminary practical findings indicate potential and highlight challenges, without claiming to be representative.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

Einleitung

Das Projekt Digital Literacy zielt darauf ab, Lehr-/Lernangebote zum KI-gestützten Erwerb akademischer Lese- und Schreibkompetenz in den Fremdsprachen Englisch, Deutsch, Französisch und Spanisch auf den Niveaustufen B2+ und C1 des Gemeinsamen Europäischen Referenzrahmens (GER) zu entwickeln. Die Angebote ermöglichen Studierenden, ihre Kompetenzen im Bereich der akademischen Textproduktion in der Zielsprache selbstgesteuert und bedarfsorientiert weiterzuentwickeln. Zugleich fördern sie AI Literacy. Auf diese Weise sollen die Studierenden auf die Herausforderungen der digitalen, mehrsprachigen Welt und insbesondere auf den Umgang mit KI vorbereitet werden.

Theoretischer Hintergrund und didaktischer Ansatz

Das Projekt wurde aufgrund der wachsenden Bedeutung von KI-Tools und deren Einfluss auf den Schreibprozess initiiert (vgl. Anson/Straume 2022: 6f.). Angegliedert an das bereits bestehende Projekt Begleitetes Autonomes Fremdsprachenlernen des Sprachlernzentrums (SLZ) verbindet der selektierte didaktische Ansatz eine lerner*innen-, handlungs- und kompetenzorientierte Nutzung von KI-Tools mit der Förderung der Lerner*innenautonomie.

Methodik

Die Plattform fobizz wurde aufgrund ihrer datenschutzkonformen Nutzung von KI-basierten Tools und Chatbots wie etwa ChatGPT für das Pilotprojekt ausgewählt. Zur Bearbeitung der Materialien und E-Learning-Einheiten sowie der Dokumentation ihres Lernprozesses nutzen die Studierenden die Lernplattform ILIAS. Auf der didaktischen Ebene wurden Lernmaterialien zur Förderung der akademischen Lese- und Schreibkompetenz mit KI-Unterstützung konzipiert. Dabei erhalten die Studierenden individuelles Feedback der KI. Generative KI-Systeme basieren jedoch auf statistischen Wahrscheinlichkeiten ohne echtes semantisches Verständnis von Didaktik oder Lernprozessen und bergen Risiken wie inhaltliche Bias oder Halluzinationen aus Trainingsdaten (vgl. Walter 2024: 11f.). Die Materialien dienen daher primär als Anregung für kritische Auseinandersetzung und menschliche Überprüfung, um Studierende zu kritischer AI Literacy zu befähigen (vgl. Walter 2024: 11; vgl. UNESCO 2024: 27). Teilweise können sie zwischen verschiedenen Aufgaben bedarfsorientiert auswählen. Dieser Ansatz trägt zur Förderung eines aktiven, selbstregulierenden und kompetenzorientierten Lernens bei. Im Laufe des Lernprozesses entwickeln die Studierenden schrittweise die Fähigkeit, die Validität und Relevanz von KI-Outputs kritisch zu bewerten. Die

systematische Dokumentation des Schreib-, Lese- und Reflexionsprozesses erfolgt in einem für diese Module neu konzipierten E-Portfolio. Die dort eingereichten Texte, Prompts, Chatverläufe inkl. der modellgenerierten Ausgaben werden von der moduldurchführenden Lehrkraft kritisch begutachtet und kommentiert. Vorteile der vorliegenden Methodik liegen in der Erfassung von Lernerfolgen sowie der Identifizierung von Verbesserungspotenzialen. Zum Semesterende wird eine schriftliche Modulabschlussprüfung ohne KI-Einsatz durchgeführt.

Umsetzung

Es finden fünf Präsenzveranstaltungen statt. Zwei davon konzentrieren sich auf die Förderung der Lernorganisation und der KI-Nutzung. Drei sprachpraktische Übungen sind der Reflexion sowie der mündlichen Sprachkompetenz gewidmet. Im Laufe des Semesters müssen die Studierenden in einem dreiwöchigen Rhythmus sechs E-Learning-Einheiten bearbeiten. Zusätzlich nutzen sie das bereitgestellte Selbstlernmaterial. Die sprachlichen Lernziele orientieren sich an den etablierten Standards des GER sowie den Vorgaben des SLZ. Bei den entwickelten Lernmaterialien liegt der Fokus auf der Planung, Gliederung und Verfassung verschiedener Textsorten wie etwa formellen E-Mails, Zusammenfassungen, Aufsätzen oder Abstracts, der Erkennung von Hauptinhalten komplexer akademischer Texte, der Erschließung stilistischer und formaler Besonderheiten akademischer Textsorten sowie der Anwendung von Lese- und Schreibstrategien im akademischen Kontext. Im Rahmen eines ersten Seminars wird den Studierenden grundlegendes Wissen im Bereich der künstlichen Intelligenz vermittelt, was der sukzessiven Entwicklung ihrer AI Literacy dient. Diese wird hier als Kompetenz verstanden, KI-Systeme zu verstehen, zu nutzen und kritisch zu hinterfragen; einschließlich ethischer und rechtlicher Implikationen (vgl. Walter 2024: 11; vgl. UNESCO 2024: 27–36). Dabei erhalten sie einen kompakten Überblick über die Funktionsweisen generativer KI, insbesondere Large Language Models. Potenziell auftretende Probleme wie etwa Halluzinationen und Bias werden zur Förderung einer kritischen Hinterfragung der KI-generierten Ergebnisse thematisiert (vgl. Walter 2024: 11f.). Auch ethische Aspekte sowie Copyright und Datenschutz werden behandelt. Das erste Seminar führt in einfache Prompting-Techniken ein. Darauf aufbauend werden im zweiten Seminar fortgeschrittene Prompting-Techniken (*chain of thought, role expert, self-consistency*) erläutert (vgl. ebd.: 15). Außerdem lernen die Studierenden Strategien, um ihre KI-Ergebnisse zu optimieren, was die Lerner*innenautonomie im Umgang mit KI-Anwendungen fördert. Die Sensibilisierung für einen reflektierten und verantwortungsvollen Umgang mit den durch KI-generierten Ergebnissen stellt einen substanziellen Aspekt der Module dar und erfolgt kontinuierlich. Im Wintersemester 2024/25 wurden die beiden Englisch-Module »Academic Reading and Writing« auf den Niveaustufen B2+ und C1 erstmalig durchgeführt und im

Sommersemester 2025 sowie Wintersemester 2025/26 fortgeführt. Ein Transfer der im Rahmen des Pilotprojekts gewonnen Erkenntnisse zur KI-Nutzung auf weitere Sprachkurse des SLZ wird angestrebt.

KI-Integration in die Fremdsprachenmodule

Der folgende Abschnitt dokumentiert die systematische Implementierung von KI-Tools in die speziell dafür entwickelten universitären Fremdsprachenmodule des SLZ. Dabei werden sowohl die praktischen Anwendungen als auch die daraus resultierenden Erfahrungen der durchführenden Lehrkraft aufgezeigt. Im Mittelpunkt steht die Integration von KI-gestützten Lernformen in die E-Learning-Einheiten eines Sprachmoduls zur Förderung akademischer Schreib- und Lesekompetenz sowie kritischer Reflexion des KI-Einsatzes. Die KI-Integration konzentriert sich primär auf die Förderung übergeordneter Schreibkompetenzen, die das Gesamtbild und die strukturelle Qualität eines Textes betreffen. Diese Higher Order Concerns (vgl. Buck 2025: 80f.) umfassen verschiedene Dimensionen des Schreibprozesses. Chatbots werden systematisch für Brainstorming-Prozesse, beispielsweise zur Entwicklung von Kernthesen eingesetzt (vgl. ebd.: 141). Dabei wurden unmodifizierte Standard-Chatbots ausgewählt, die auf der fobizz-Plattform zugänglich sind. Der Ansatz ermöglicht es Studierenden, ihre Gedanken zu strukturieren und neue Perspektiven zu entwickeln, bevor sie mit dem eigentlichen Schreibprozess beginnen. Im Rahmen der Erstellung argumentativer Texte werden sowohl eigene als auch KI-generierte Argumente einer kritischen Prüfung unterzogen und in kooperativen Prozessen iterativ optimiert. Darüber hinaus generieren die Bots zusätzliche Argumente sowie Gegenargumente, wodurch die argumentative Tiefe und Ausgewogenheit der Texte gefördert werden (vgl. Buck 2025: 201). Die KI fungiert zudem als Unterstützung bei der Textstrukturierung, bewertet bestehende Strukturen und unterbreitet konkrete Verbesserungsvorschläge (vgl. ebd.: 205–208). Hervorzuheben ist die Implementierung sokratischer Dialoge, in denen Studierende durch gezielte Fragen zu tieferer Reflexion über ihre Texte angeleitet werden (vgl. ebd.: 136f.). Der Einsatz von KI transformiert Prozesse akademischer Textproduktion und Wissensvermittlung. Große Sprachmodelle basieren jedoch auf statistischen Wahrscheinlichkeitsmodellen, die sprachliche Daten in probabilistischer Form kombinieren, ohne semantisches Verständnis oder Einsicht in fachliche und didaktische Zusammenhänge zu besitzen. Ihre Vorschläge repräsentieren Wahrscheinlichkeiten sprachlicher Muster statt inhaltlicher Validität und können Verzerrungen des Trainingsmaterials übernehmen (vgl. Bender et al. 2021: 616f.). Die KI-generierten Outputs sind daher als potenziell nützliche, aber grundsätzlich fehlbare Entwürfe zu betrachten, die eine kritische menschliche Prüfung erfordern. Das Lehrkonzept sieht die gezielte Arbeit mit KI-Ergebnissen in Refle-

xionsprozessen vor. Dies geschieht einerseits in der Gruppe in den Präsenzphasen und wird insbesondere im Rahmen der E-Learning-Einheiten von den Studierenden individuell im E-Portfolio dokumentiert. Studierende analysieren, bewerten und modifizieren die generierten Vorschläge, um eine kritische und lernzielorientierte Nutzung zu entwickeln. Die im E-Portfolio dokumentierten Reflexionen und Überarbeitungen werden anschließend von der Lehrkraft nochmals kritisch geprüft und kommentiert, um sicherzustellen, dass mögliche Fehlschlüsse der KI-Ergebnisse erkannt und didaktisch aufgearbeitet werden. Diese Praxis fördert eine kritische AI Literacy – die Fähigkeit, maschinell erzeugte Inhalte differenziert zu bewerten und verantwortungsvoll in den eigenen Lern- und Schreibprozess zu integrieren. Auf der technisch-formalen Ebene fokussieren sich die Lower Order Concerns (vgl. Buck 2025: 78f.) auf grundlegende Aspekte der Satzebene und dienen der Optimierung selbst verfasster Texte während und nach dem Schreibprozess. Die KI-Analyse evaluiert Redemittel und verwendetes (Fach-)Vokabular. Diese systematische Bewertung unterstützt Studierende dabei, ihre sprachlichen Mittel textsortenspezifisch einzusetzen. Grammatische Phänomene wie Satzstrukturen sowie Orthografie und Interpunktion werden analysiert und korrigiert (vgl. ebd.: 212f.). Anhand der von der KI erhaltenen Korrekturen können Studierende ihre gemachten Fehler und Schwachstellen im verfassten Text schnell identifizieren, sie reflektieren und korrigieren. Die vorliegende Funktionalität gestattet eine gezielte Verbesserung der formalen Textqualität. Die Teilnehmenden werden zudem regelmäßig dazu aufgefordert, Chatbots zur Erstellung interaktiver Grammatikübungen, zur Einübung von Vokabeln und Redemitteln sowie zur Erstellung weiterer Lernmittel, etwa Flashcards, einzusetzen (vgl. Poleshchuk/Schatz 2024: 9–13). Solche personalisierten Übungen unterstützen die angestrebte Binnendifferenzierung, da Studierende gemäß ihren individuellen Bedürfnissen lernen, wiederholen oder vertiefen können. Das Lehrkonzept in den E-Learning-Einheiten basiert auf einer systematischen Prompting-Progression, die von angeleiteten Beispielprompts zu zunehmender Autonomie der Studierenden führt (vgl. Knellesen 2025). Ziel ist es, eine Prompting-Kompetenz aufzubauen, die es den Studierenden ermöglicht, selbstständig effektive Prompts zu entwickeln und anzuwenden.

Evaluationsergebnisse

Die Begleitforschung umfasst bei entsprechend vorliegender Einwilligung die anonymisierte Auswertung der bearbeiteten E-Learning-Aufgaben und Klausurtexte (Ergebnisse ausstehend). In der ersten Sitzung werden die Teilnehmenden (TN) befragt, wie häufig sie wozu (z.B. Vorbereitung, Formulierung oder Korrektur von Texten) KI-Tools nutzen. Zudem werden alle TN (N = 13, n = 6 aus WS 24/25, n = 7 aus SS 25) zum Modulabschluss gebeten, einen Evaluationsfragebogen auszufül-

len. Aufgrund der geringen TN-Zahl sind die Ergebnisse jedoch nicht repräsentativ. Der für die Bewertung der KI-Nutzung relevante Teil besteht aus Likert- und offenen Fragen. In ihrer Gesamtheit wurde das Modul als durchweg positiv bewertet (6 = eher zufrieden, 7 = sehr zufrieden). Die TN attestierten den KI-Tools eine hohe Nützlichkeit für ihren individuellen Lernprozess (1 = trifft eher nicht zu, 7 = trifft eher zu, 5 = trifft zu). Zudem lässt sich feststellen, dass die KI-Nutzung der TN im Vergleich zu Beginn des Moduls zugenommen hat. Basierend auf einzelnen Antworten und Kommentaren des offenen Frageteils sowie mündlichen Rückmeldungen der Studierenden werden nachfolgend weitere Ergebnisse abgeleitet. Sie haben somit den Charakter einer anekdotischen Evidenz und sind nicht repräsentativ. Es zeigten sich spezifische Problemfelder in der praktischen Anwendung. Einige TN reagierten frustriert, wenn die KI nicht den gewünschten Output generierte oder zu ausführlich antwortete, obwohl um eine kurze Antwort gebeten wurde. Problematisch erwies sich, dass die KI bisweilen Informationen ohne explizite Aufforderung generierte oder Textstellen eigenständig bearbeitete und korrigierte. In den Übungsaufgaben wurden häufig die Lösungen durch den Chatbot bereitgestellt, bevor die Studierenden die Möglichkeit erhielten, ihre eigenen Antworten zu entwickeln oder zu korrigieren. Die Rückmeldungen einiger Studierender zeigten, dass sie befürchten, ihre individuelle Kreativität sowie die intellektuelle Kontrolle über ihre selbst produzierten Texte zu verlieren. Es wurde die Sorge geäußert, dass infolge kontinuierlicher Textrevision unter Zuhilfenahme von KI-Technologien nur noch eine marginale Korrelation zum ursprünglichen Text besteht. Der adäquate Umgang mit diesen Herausforderungen erfordert von den Studierenden ein hohes Maß an Selbstdisziplin und Selbstreflexion.

Fazit

Die Ergebnisse der begleitenden Evaluation weisen auf eine grundsätzlich positive Akzeptanz von KI-Tools im Lernkontext hin, wenngleich die geringe Stichprobe und anekdotische Evidenz die Aussagekraft einschränken. Für künftige empirische Untersuchungen sind größere und heterogenere Stichproben erforderlich, um differenzierte Aussagen zu Nutzungsintensität, Lernerfolg und Kompetenzentwicklung treffen zu können. Darüber hinaus sollte die qualitative Dimension der KI-Interaktion – etwa Frustrationserlebnisse, Kontrollverlust oder wahrgenommene Beeinträchtigung der Kreativität – vertieft untersucht werden. Die identifizierten Herausforderungen verdeutlichen den Bedarf an Lernarrangements, die Selbstreflexion, kritisches Denken und eigenständige Textproduktion fördern. Die bisher im Projekt entwickelten Konzepte bieten ein solides Fundament für die weitere Expansion KI-gestützter Lernformen im Hochschulbereich, wobei die Balance zwi-

schen technologischer Innovation und pädagogischer Qualität stets gewährleistet werden muss.

Literatur

Anson, Chris M./Straume, Ingerid S. (2022): »Amazement and Trepidation. Implications of AI-Based Natural Language Production for the Teaching of Writing«, in: Journal of Academic Writing 12 (1), S. 1–9. https://doi.org/10.18552/joaw.v12i1.820

Bender, Emily M./Gebru, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret (2021): »On the Dangers of Stochastic Parrots. Can Language Models Be Too Big?«, in: FaccT 21. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, S. 610–623.

Buck, Isabella (2025): Wissenschaftliches Schreiben mit KI (= Studieren, aber richtig, Band 23), Tübingen: UVK Verlag.

Knellesen, Jennifer (2025): Aufbau der KI-Nutzungskompetenz. siehe: https://de.linkedin.com/posts/jennifer-knellesen-b0084714_kikompehenz-bildungmitki-künstlicheintelligenz-activity-7324366341807763457-E796

Poleshchuk, Alisa/Schatz, Thomas (2024): Fremdsprachenunterricht mit KI gestalten. Eine Handreichung für die Praxis (= Harzer Hochschultexte, Band 16), Wernigerode.

United Nations Educational, Scientific and Cultural Organization (2024): AI competency framework for students, Paris.

Walter, Yoshija (2024): »Embracing the future of Artificial Intelligence in the classroom. The relevance of AI literacy, prompt engineering, and critical thinking in modern education.«, in: International Journal of Educational Technology in Higher Education 21 (15), S. 1–29.

Fostering AI Literacy in Higher Education

The Role of Students' Usage Patterns[1]

Sittipan Yotyodying[2], Maria Zirenko[3], Sabine Fabriz[4] und Holger Horz[5]

The expanding use of AI in education has driven governmental initiatives underlining the need for critical, ethical, and effective AI use. The TAM model suggests that more AI use could improve students' AI literacy, and evidence shows that promoting AI literacy among university students benefits their well-being and learning outcomes. However, research on key preconditions of AI literacy remains unclear. Moreover, there is a need to distinguish between self-reported and actual skills. In a survey (N = 247), we found that three frequent AI usage patterns (write essays, prepare presentations, organize workflows) were linked to certain dimensions of self-reported AI literacy and to specific facets of a knowledge test. Academic uses of AI may foster confidence, while not necessarily enhancing real proficiency. Results suggest the way students use AI matters more than how often.

Förderung von AI Literacy in der Hochschulbildung.
Die Rolle studentischer Nutzungsmuster

Die zunehmende Nutzung von KI im Hochschulbereich hat politische Initiativen hervorgebracht, die einen kritischen, ethischen und effektiven Einsatz fordern. Das TAM-Modell legt nahe, dass schon die intensive Nutzung von KI die AI Literacy von Studierenden fördern könnte. Empirische Befunde deuten darauf hin, dass AI Literacy mit Wohlbefinden und besseren Lernerfolgen einhergeht. Welche Prädiktoren dabei eine Rolle spielen, ist unklar. Außerdem ist es notwendig, selbstberichtete Kompetenzen von tatsächlichen Fähigkeiten zu unterscheiden. In einer Befragung (N = 247) zeigte sich, dass drei Nutzungsmuster (Essays verfassen, Präsentationen vorbereiten, Arbeitsabläufe organisieren) mit Dimensionen von AI Literacy im Selbstbericht als auch mit Facetten eines Wissenstests zusammenhingen. Die

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000–0002-1225-3107
3 ORCID: 0000–0003-4495-4220
4 ORCID: 0000–0003-2262-9283
5 ORCID: 0000-0002-5173-0252

Ergebnisse deuten darauf hin, dass die Qualität der Nutzung wichtiger ist als die Häufigkeit.

Background

Relevance in education and in higher education

The development of artificial intelligence (AI) has accelerated rapidly, leading to its widespread integration into education, from elementary schools to universities and vocational training. This expansion has strengthened the importance of AI literacy, i.e., the knowledge and ability to use AI effectively, critically, ethically, and responsibly (Ng/Leung/Chu/Qiao 2021). Policymakers have explicitly emphasized this need as well. For example, the EU AI Act requires all actors involved in operating and using AI systems (European Commission 2025). In Germany, both the Federal Ministry of Education and Research (BMBF) and the Standing Conference of the Ministers of Education and Cultural Affairs (KMK) have published guidelines since 2023 that highlight the need for AI literacy in school and higher education contexts (BMBF 2023; KMK 2024). This policy emphasis underscores the necessity of equipping educators and students with competencies to handle both opportunities and risks of AI. Beyond policy, international organizations such as UNESCO IITE and SOU (2023) have similarly called for AI literacy as a cornerstone of digital citizenship, arguing that the ability to engage critically with AI is a prerequisite for democratic participation and social inclusion in increasingly digital societies.

AI Literacy

Past research has broadly defined AI literacy as the set of knowledge, skills, and attitudes that enable individuals to interact with AI systems in an effective, critical, and ethical manner (e.g., Carolus et al. 2023; Long/Magerko 2020). Additionally, scholars have emphasized that AI literacy should be viewed not only as a technical competence but also as a socio-technical and ethical capacity, encompassing awareness of issues such as bias, fairness, and accountability in AI systems (Tuomi/Cachia/Villar-Onrubia 2023). From an educational policy perspective, the OECD AI Literacy Framework (AILit) postulates that AI literacy includes technical understanding (e.g., how AI systems process data), the ability to use AI tools responsibly, and the capacity to evaluate AI outputs critically (OECD 2025). Importantly, AI literacy also requires practical competencies such as prompt engineering to improve the accuracy and usefulness of AI responses, as well as the skill to identify and mitigate hallucinations, i.e., AI-generated but factually incorrect information (Kasneci et al. 2023).

Assessment of AI Literacy

In recent years, researchers have proposed a wide range of instruments to assess AI literacy (for an overview, see the meta-analysis by Lintner 2024). These assessments typically cover dimensions such as technological knowledge, ethical considerations, problem-solving abilities, and the use of AI applications. Most existing instruments rely on self-reported skills, with only a few including performance-based tests. Lintner et al. (2024) identified thirteen self-report scales and three performance-based measures. This imbalance is problematic, as self-reports often overestimate competencies. Indeed, prior research found a negative connection between AI tool usage and critical thinking skills (Gerlich 2025), suggesting that frequent AI use may boost confidence without necessarily improving proficiency. Consequently, distinguishing between perceived (self-reported) abilities and actual abilities is essential for valid assessment. Furthermore, the sample for which an instrument has been validated must be taken into account, differentiating between university students, K-12 students, teachers, or the general population.

Antecedents and Consequences of AI Literacy

Promoting AI literacy has been associated with positive outcomes for university students, including enhanced well-being and improved academic performance (Xiao et al. 2024). Yet, the conditions that foster AI literacy are not fully understood. The Technology Acceptance Model (TAM; Davis 1989) suggests that frequent use of technology enhances related skills, implying that more AI usage could naturally improve AI literacy. However, evidence is mixed. The International Computer and Information Literacy Study (ICILS 2023) found that years of computer experience correlated positively with ICT literacy, whereas academic-media multitasking was negatively associated (Fraillon 2024). These findings indicate that how technology is used matters more than how often it is used. Similarly, in higher education, little is known about the usage patterns students follow when using AI tools for academic purposes and how these relate to AI literacy.

Aims

The present study aimed to (a) explore usage patterns of AI tools in study-related activities among university students and (b) examine how these patterns are connected to AI literacy, distinguishing between self-reported abilities and actual knowledge. By integrating insights from prior research with new empirical evidence, this work contributes to understanding how AI use in higher education shapes students' competencies for responsible and effective AI engagement.

Methods

Sample & Procedure

Analyses in the present study were based on data from the project »Studis-KI« (Studis-KI 2025). An online survey using SoSciSurvey was conducted. Data collection took place between November 2024 and August 2025. Ethical approval was obtained prior to the start of the survey. A total of $N = 247$ university students (80 % female, $M_{age} = 23.06$, $SD = 4.39$) who indicated using AI for study purposes participated in this survey. Sixty-two percent of students had a background in psychology (64 % Bachelor, 11 % Master) and 23 % in teacher education. Participants perceived either participant minutes ($n = 149$) or a voucher ($n = 98$).

Measures

In this section, we describe the measurement instruments used to assess students' study-related experiences with AI tools, as well as different measures of AI literacy, including both self-report scales and a knowledge test. Where applicable, we used Cronbach's alpha to estimate internal consistency. An alpha value of ≥ 0.70 is generally considered satisfactory (Tavakol/Dennick 2011).

AI Literacy (Self-report)

Self-reported AI literacy was assessed using four subscales of the META AI Literacy Scale (MAILS; Carolus et al. 2023), with an 11-point rating scale. The original MAILS subscales focus on AI use in a general context. In the present study, we specifically examined AI use in higher education. Therefore, we adapted the items to fit the higher education context, for example, by adding phrases such as »my studies« where applicable (e.g., »I can use AI tools to make my studies more efficient.«). The four subscales included Use and Apply (4 items; $\alpha = .91$), Know and Understand (4 items; $\alpha = .69$), Detect AI (4 items; $\alpha = .78$), and AI Ethics (4 items; $\alpha = .77$). It must be noted that the internal consistency of the Know and Understand subscale was below .70. Item–total correlations suggested deleting one item, which would have improved the internal consistency to .72. However, to preserve content validity, we decided not to delete this item, as the subscale has previously been shown to be reliable in a large sample.

AI Literacy (knowledge test)

For an objective measure of AI literacy, we developed a short knowledge test consisting of six single-choice items with four answer options for critical AI use based on recommendations of Tancredi (2024). We developed each item for six of the seven proposed strategies; the strategy »implement retrieval-augmented generation« was not addressed because it is less applicable in students' everyday use. Each item had one correct answer. All items had item difficulty indices (*p*) that fell within the recommended range: (a) increase awareness/attention (.79), (b) use a more advanced model (.48), (c) provide explicit instruction (.89), (d) provide sample answers (.81), (e) provide full context (.31), and (f) verify outputs (.84).[6]

Quantity of AI Use for Study-Related Activities (Usage Patterns, UP)

As a proxy for the quantity of AI use in study-related activities (see Table 1), we asked students to rate how often they engaged in 14 types of AI-supported study practices: »*How often do you use AI tools for the following study-related activities?*« Responses were given on a 5-point scale (1 = *never*, 5 = *every day*). The scale was adapted from an instrument originally developed to assess students' use of ICT for study purposes in ICILS 2018 (Vennemann et al. 2021). For this study, the items were revised to reflect AI-specific applications rather than general ICT use.

Background Variables

Students were asked to provide demographic information (e.g., gender, age, study program) and self-reported data on their Abitur GPA and the frequency of use of 51 different AI tools. In addition, they were asked to indicate their prior knowledge (yes/no) in 13 different areas of AI (e.g., machine learning, deep learning, neural networks) or to state that they had no prior knowledge of the topic.

Results

Descriptives and Correlational Analyses

Descriptive analyses showed that, among 51 AI tools, most students reported frequently using ChatGPT (92 %), followed by DeepL (30 %), and Canva (15 %), respectively. Almost half of the students (46 %) reported having no knowledge about AI concepts, while about a third reported some understanding of advanced AI topics: 33 %

6 The scale is available from the first author for research purposes upon request.

machine learning, 27 % deep learning, and 26 % neural networks. Correlational analyses revealed positive associations among the dimensions of self-reported AI literacy (AIL). Guided by prior research (Gerlich 2025) that self-perceived AI competencies and actual knowledge do not necessarily align, this study did not aim to identify latent AI literacy constructs but rather to examine how predefined and conceptually distinct dimensions of self-reported AIL and knowledge-based performance relate to one another. Consistent with these assumptions, the dimensions of self-reported AIL and those of the knowledge test were largely uncorrelated (see Table 1).

Table 1: Correlations between Dimensions of AI Literacy (AIL): Self-Report versus Knowledge Test

AIL Measures	*N*	*M*	*SD*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Self-Report													
(1) Use & Apply	243	8.45	1.83	—									
(2) Know & Understand	243	7.80	1.64	.67**	—								
(3) Detect AI	243	6.33	2.05	.40**	.59**	—							
(4) AI Ethics	243	7.36	1.90	.42**	.70**	.57**	—						
Knowledge Test													
(5) Increase Awareness	241	0.79	0.41	.01	.10	.02	.04	—					
(6) Advanced Model	241	0.48	0.50	.01	.00	.04	.03	.02	—				
(7) Explicit Instruction	241	0.89	0.31	.17**	.20**	.01	.14*	.25**	.07	—			
(8) Example Answers	241	0.81	0.39	.00	.04	.02	.05	.17**	.05	.14*	—		
(9) Full Context	241	0.32	0.47	.07	.08	.14*	.14*	.01	.03	.12	.10	—	
(10) Verify Outputs	241	0.84	0.37	.13*	.11	.04	.08	.51**	-.06	.33**	.25**	.10	—

Note: *$p < .05$. **$p < .01$.

Exploring Usage Patterns of AI Tools For Study Purposes Among University Students

Table 2 shows descriptive statistics for fourteen AI usage patterns among university students. The results showed that students used AI most often for research and

information gathering ($M = 3.47$, $SD = 1.17$) and for preparing research papers and essays ($M = 3.16$, $SD = 1.14$), whereas AI tools were rarely employed for creative tasks such as video content ($M = 1.48$, $SD = 0.93$) or audio production ($M = 1.47$, $SD = 0.93$).

Table 2: Descriptive Statistics for AI Usage Patterns among University Students

AI Usage Patterns (1 = *never*, 5 = *every day*)	***N***	***M***	***SD***
Research and information gathering on the internet*	247	3.47	1.17
Preparation of research papers and essays*	247	3.16	1.14
Working on worksheets and exercises*	247	3.06	1.27
Preparation of presentations	247	2.67	1.26
Working on exams and quizzes	247	2.35	1.33
Research and preparation of information for long-term projects	247	2.31	1.24
Use of software or applications to develop specific skills or for a subject (e.g., math tutoring)	247	2.30	1.32
Structuring of information, e.g., as lists, tables, or mind maps	247	2.23	1.25
Use of programming environments to work on tasks (e.g., programming projects)	247	2.21	1.35
Organization of one's own time and work processes**	247	2.19	1.32
Analysis and presentation of data, e.g., in charts and graphs	247	2.06	1.24
Working online with other students	247	2.00	1.19
Creation of videos to explain study content	247	1.48	1.02
Creation of video or audio productions	247	1.47	0.93

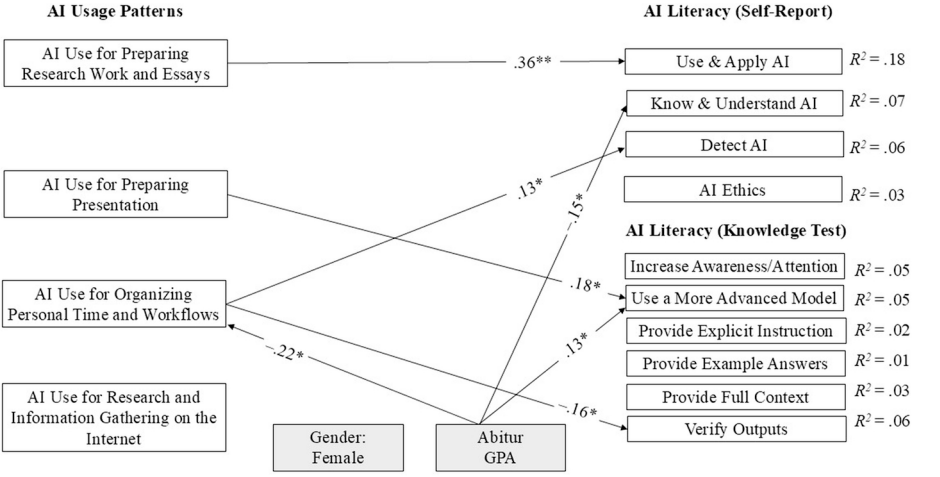
Note: * patterns with the highest frequencies selected for analyses; **pattern selected for analyses

AI Usage Patterns Associated with AI Literacy Among University Students

Among the 14 AI usage patterns, we selected four as predictors of AI literacy. We chose *AI use for research and information gathering*, *AI use for preparation of research papers and essays*, and *AI use for preparation of presentations*, as the most frequently reported activities among the 14 patterns. *AI use for working on worksheets and exercises* ranked third in frequency, but was not included because it substantially overlaps conceptually with *AI use for preparation of research papers and essays* and thus could introduce redundancy in the analysis. Instead, we selected *AI use for organization of one's own time and work processes*, not because of its frequency of use, but due to its conceptual rel-

evance for self-regulated learning skills (e.g., Zimmerman 2002), which are central to understanding students' AI literacy.

Figure 1: Path Model for AI Usage Patterns Associated with AI Literacy Among University Students



Note. $N = 247$. $^*p < .05$. $^{**}p < .01$.
For simplicity, correlations among usage patterns and among AI literacy indicators were estimated but are not displayed.

A path analysis was conducted using Mplus 8.8 (Muthén & Muthén 1998–2022) to examine the hypothesized associations while controlling for gender and Abitur GPA in a sample of 247 participants. It must be noted that Abitur GPA was included as a control variable to account for general academic achievement, thereby ensuring that the associations between AI usage patterns and AI literacy are examined independently of prior educational performance.

All study variables were specified as manifest variables; grouping labels are included solely for conceptual clarity. Model fit indices indicated a mixed fit, $\chi^2(2, N = 247) = 19.17$, CFI = .98, SRMR = .03, RMSEA = .19. Overall, model fit was acceptable, with strong support from CFI and SRMR; although RMSEA indicated poor fit, this index is unreliable in low-df models (Kenny et al., 2015), so findings should be interpreted with caution. Three of the four AI usage patterns (UP, reported in Table 1) were significantly related to certain dimensions of self-reported AIL as well as some knowledge test dimensions: Regarding self-reported AIL, the UP *AI use for preparing research papers and essays* was positively associated with the perceived ability to use and apply AI, while the UP *AI use for organizing workflows* was positively associated with the perceived ability to detect AI. No significant associations for UPs were found with perceived knowledge and understanding of AI or with perceived AI ethics. Interestingly, Abitur GPA was negatively associated with perceived

knowledge and understanding of AI. Likewise, it was negatively associated with the UP *AI use for organizing workflows*. The UP *AI use for research and information gathering* showed no significant associations with any literacy dimensions. For the knowledge test items, the UP *AI use for preparing presentations* and Abitur GPA were positively associated with knowledge of recognizing when an advanced model is needed. In contrast, the UP *AI use for organizing workflows* was negatively associated with knowledge of validating AI output. Gender was unrelated to both AI usage and literacy dimensions.

Discussion

Summary & Discussion of Results

The study identified four patterns of AI use in higher education. Three were most frequently employed among students: using AI tools to research and information gathering on the internet, to prepare research papers or essays, and to prepare presentations. The fourth pattern is relevant to self-regulated learning skills as it involves organizing personal time and workflows. Three of these patterns were found to be associated with specific dimensions of AI literacy, both self-reported (e.g., Use and Apply AI, and Detect AI) and knowledge test (e.g., Use a More Advanced Model, and Verify Outputs). Notably, AI use for information gathering showed no connection to either self-reported or tested AI literacy. Furthermore, a negative relationship was observed between Abitur GPA and the pattern of using AI for organizing personal time and workflows, as well as knowledge and understanding of AI.

In the three patterns that showed associations, students interact more actively with the AI tools and typically require multiple feedback loops, which in turn could explain why they are linked to AI literacy. In contrast, using AI for quick information searches involves less active engagement and critical evaluation. Interestingly, we found students with lower Abitur GPA appeared to rely more on AI, while those with higher Abitur GPA report being more critical in their AI usage. This interpretation aligns with prior research showing that students' engagement with generative AI is heterogeneous and task-dependent, ranging from low overall reliance to extensive use for assignment-related activities (Stojanov et al. 2024)

The findings suggest that while academic uses of AI may foster confidence in handling these tools, such usage does not automatically translate into actual proficiency. This highlights the crucial point that *how* students use AI matters more than *how often* they use it. Moreover, the distinction between self-reported and tested AI literacy becomes essential in both research and practice: confidence alone cannot be equated with competence. Designing curricula and interventions should therefore aim at promoting measurable skills, not merely self-perceptions.

Limitations

The study sample consisted mainly of female students and students of psychology, which limits the generalizability of the results. In addition, the strong focus on the Goethe University Frankfurt context (93 % of participants) limits the transferability to other institutions or subject areas. It must be noted that the working model consisted of many study variables. Increasing the sample size would allow us to gain information about the robustness of the model. Finally, the knowledge test requires further development, due to its limited number of items to measure each strategy for critical AI evaluation skill and some items that were rather easy. Strategies to increase the test's reliability and depth include particularly the inclusion of more items per strategy and an in-depth analysis of the quality of distractors.

Further Investigation of the Dataset

Theoretical frameworks from psychology will be employed to explain antecedents and consequences of AI literacy. Ongoing work will therefore examine additional predictors of AI literacy, including further usage patterns, personality traits (e.g., Paulhus/Williams 2002), and motivational beliefs (e.g., self-determination theory, Deci/Ryan 2000). Moreover, future studies will extend the focus to other groups, such as school students and the elderly. In addition to antecedents, consequences of AI literacy will be explored in relation to achievement, well-being, and mental health in future analyses based on the Studis-KI dataset.

Practical Implications & Outlook

The findings provide empirical evidence to inform the development of interventions that foster AI literacy in higher education. For students, guidelines should focus on activities and usage patterns that strengthen critical engagement with AI and foster competence beyond surface-level interaction. For university lecturers, the study underscores that the use of AI is not homogeneous among students; targeted training and clear guidelines are necessary to support reflective and responsible use. This could be reflected in adaptation of curriculum designs (e.g., seminars, study programs, workshops) to systematically strengthen critical AI skills, or adaptation of assessment strategies to fit the new curricula (Digital Education Council 2025). To date, the link from AI literacy to student success (e.g., achievement, motivation, well-being) is still underexplored. A meta-analysis (e.g., Deng et al. 2024) demonstrated that the use of AI in education was connected with diverse educational outcomes. However, this connection should be further investigated systematically.

Scientific Significance of the Study

This study advances the understanding of how specific AI usage patterns relate to AI literacy, demonstrating that certain types of use can foster competence while others may undermine critical evaluation. For educators, the findings provide valuable guidance on integrating AI tools into teaching and learning, enabling targeted support to address both opportunities and challenges. Moreover, the insights gained can serve as a foundation for developing seminar and workshop concepts that strengthen AI literacy. By highlighting how diverse patterns of AI use contribute to academic activities, the study underscores the importance of promoting responsible and informed engagement with AI tools in higher education.

References

Bundesministerium für Bildung und Forschung (2023): BMBF-Aktionsplan. Künstliche Intelligenz. Neue Herausforderungen chancenorientiert angehen, https://www.bmbf.de/SharedDocs/Downloads/de/2023/230823-executive-summary-ki-aktionsplan.pdf?__blob=publicationFile&v=1 vom 21.09.2025.

Carolus, Astrid/Koch, Martin J./Straka, Samantha/Latoschik, Marc E./Wienrich, Carolin (2023): »MAILS – Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies«, in: Computers in Human Behavior: Artificial Humans 1, S. 100014.

Davis, Fred D. (1989): »Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology«, in: MIS Quarterly 13, S. 319.

Deci, Edward L./Ryan, Richard M. (2000): »The »What« and »Why« of goal pursuits: human needs and the self-determination of behavior«, in: Psychological Inquiry 11, S. 227-268.

Deng, Ruiqi/Jiang, Maoli/Yu, Xinlu et al. (2025): »Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies«, in: Computers & Education 227, S. 105224.

Digital Education Council (2025): The next era of assessment. A global review of ai in assessment design, https://www.digitaleducationcouncil.com/post/the-next-era-of-assessment-a-global-review-of-ai-in-assessment-design vom 07.07.2025.

European Commission (2025): AI Act, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#: :text=The%20AI%20Act%20is%20the,legal%20framework%20on%20AI%20worldwide vom 21.09.2025.

Fraillon, Julian (Ed.) (2024): An International Perspective on Digital Literacy: Results from ICILS 2023, International Association for the Evaluation of Educational Achievement (IEA).

Gerlich, Michael (2025): »AI tools in society: impacts on cognitive offloading and the future of critical thinking«, in: Societies 15, S. 6.

Kasneci, Enkelejda/Sessler, Kathrin/Küchemann, Stefan/Bannert, Maria/Dementieva, Daryna/Fischer, Frank/Gasser, Urs/Groh, Georg/Günnemann, Stephan/Hüllermeier, Eyke/Krusche, Stephan/Kutyniok, Gitta/Michaeli, Tilman/Nerdel, Claudia/Pfeffer, Jürgen/Poquet, Oleksandra/Sailer, Michael/Schmidt, Albrecht/Seidel, Tina/Stadler, Matthias/Weller, Jochen/Kuhn, Jochen/Kasneci, Gjergji (2023): »ChatGPT for good? On opportunities and challenges of large language models for education«, in: Learning and Individual Differences 103, S. 102274.

Kenny, David A./Kaniskan, Bora/McCoach, D. Betsy (2015): »The performance of RMSEA in models with small degrees of freedom«, in: Sociological Methods & Research 44(3), S. 486–507

Kultusministerkonferenz (2024): Handlungsempfehlung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen, https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2024/2024_10_10-Handlungsempfehlung-KI.pd vom 21.09.2025.

Lintner, Tomáš (2024): »A systematic review of AI literacy scales«, in: npj Science of Learning 9.

Long, Duri/Magerko, Brian (2020): »What is AI Literacy? Competencies and design considerations«, in: Regina Bernhaupt/Florian Mueller/David Verweij et al. (Hg.), Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, New York, NY, USA: ACM, S. 1-16.

Ng, Davy T. K./Leung, Jac K. L./Chu, Samuel K. W./Qiao, Maggie (2021): »Conceptualizing AI literacy: An exploratory review«, in: Muthén, Linda K./Muthén, Bengt O. (1998–2022): Mplus User's Guide. Version 8.8, Los Angeles, CA: Muthén & Muthén, in: Computers and Education: Artificial Intelligence 2, S. 100041.

Paulhus, Delroy L./Williams, Kevin M. (2002): »The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy«, in: Journal of Research in Personality 36, S. 556-563.

Stojanov, Ana/Liu, Qian/Koh, Joyce Hwee Ling (2024): »University students' self-reported reliance on ChatGPT for learning: A latent profile analysis«, in: Computers & Education: Artificial Intelligence 6, S. 100243.

Studis-KI (2025): Project Studis-KI: Students in the Age of AI: Antecedents, consequences, and ethical perspectives of AI usage and AI literacy. Research Team: Yotyodying, S., Zirenko, M., Fabriz, S., Ullrich, M., & Horz, H., Department of Educational Psychology, Goethe University Frankfurt.

Tancredi, Lucy (2024): AI strategies series: 7 ways to overcome hallucinations, https://insight.factset.com/ai-strategies-series-7-ways-to-overcome-hallucinations vom 21.09.2025.

Tavakol, Mohsen/Dennick, Reg (2011): »Making sense of Cronbach's alpha«, in: International Journal of Medical Education 2, S. 53-55.

Tuomi, Ilkka/Cachia, Romina/Villar-Onrubia, Daniel (2023): On the futures of technology in education – Emerging trends and policy implications, Publications Office of the European Union.

UNESCO (2023): Global practices evaluation & assessment toolkit: Advancing artificial intelligence-supported global digital citizenship education, Moscow, https://iite.unesco.org/wp-content/uploads/2023/03/UNESCO-IITE-SOU.-2023.-Global-Practices-Evaluation-Assessment-Toolkit.-Advancing-Artificial-Intelligence-Supported-Global-Digital-Citizenship-Education.pdf vom 21.09.2025.

Vennemann, Mario/Eickelmann, Birgit/Labusch, Amelie/Drossel, Kerstin (2021): ICILS 2018 #Deutschland. Dokumentation der Erhebungsinstrumente der zweiten Computer and Information Literacy Study, Waxmann.

Xiao, Jingyu/Alibakhshi, Goudarz/Zamanpour, Alireza/Zarei, Mohamad A./Sherafat, Shapour/Behzadpoor, Seyyed-Fouad (2024): »How AI Literacy Affects Students' Educational Attainment in Online Learning: Testing a Structural Equation Model in Higher Education Context«, in: The International Review of Research in Open and Distributed Learning 25, S. 179-198.

Zimmerman, Barry J. (2002): »Becoming a Self-Regulated Learner: An Overview«, in: Theory Into Practice 41, S. 64-70.

KIViPro – KI-unterstützte Videoproduktion fördern und lernen[1]

Hendrik Bunke[2], Margret Plank[3], Jens Kösters[4], Matti Stöhr[5] und Zoë Kittelmann[6]

Wie können Lehr- und Lernvideos effektiv produziert und KI dabei sinnvoll eingesetzt werden? Das Projekt KIViPro am TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften entwickelt ein modulares, als OER veröffentlichtes Lernangebot für Hochschulen und Schulen, das grundlegende Konzepte, praxisnahe Handreichungen, Tool-Tipps und Best-Practice-Beispiele didaktisch strukturiert bereitstellt. Umfrageergebnisse mit über 100 Teilnehmenden zeigen, dass selbst Lehrende, die bereits über viel Erfahrung in der Videoproduktion verfügen, bislang kaum KI-gestützte Verfahren nutzen und hohen Unterstützungsbedarf äußern. Der auf der Moodle-basierten LMS-Plattform der TIB angebotene Kurs ermöglicht den Erwerb vielfältiger Basiskompetenzen und kann durch modulares Design, Community-Einbindung und KI-Integration an die rapide technische Entwicklung angepasst werden.

KIViPro – Advancing and Learning AI-Supported Video Production

How can teaching and learning videos be produced effectively and AI be used sensibly in the process? The KIViPro project at TIB – Leibniz Information Centre for Science and Technology is developing a modular learning program for universities and schools, published as an Open Educational Ressource (OER), which teaches basic concepts, practical tips, tool tips, and best practice examples. Survey results from over 100 participants indicate that even instructors with extensive experience in video production have so far made little use of AI-assisted methods and report a substantial need for support. The course offered on TIB's Moodle-based LMS platform enables the acquisition of a diverse range of foundational competencies and can be adapted to rapid technological developments through its modular design, community engagement, and integration of AI.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000–0002-1516-2382
3 ORCID: 0000–0001-8941-7563
4 ORCID: 0000–0002-3680-2086
5 ORCID: 0000–0002-2874-4832
6 ORCID: 0009–0009-0766-863X

Einleitung

Mit der Einführung von ChatGPT durch OpenAI Ende 2022 sowie kurzfristig nachfolgend weiterer ähnlich leistungsfähiger Dienste haben sich die Informationsbeschaffung, -verarbeitung und -verbreitung für viele, auch wissenschaftliche Bereiche grundlegend verändert. Künstliche Intelligenz (KI) hat einen immer stärkeren Einfluss auf die Gesellschaft und letztlich auch auf den Umgang mit Medien. Laut einer quantitativen Studie nutzen bereits rund 92 % der befragten Studierenden generative KI für das Studium (von Garrel/Mayer 2025). Durch Unterstützung von KI ist es ebenfalls möglich, Medien schneller und qualitativ hochwertiger zu produzieren, so zum Beispiel Lehr- und Lernvideos in Hochschulen und Schulen. Videos sind zu einer wichtigen Lernressource geworden und können, je nach Qualität ihrer didaktischen und technischen Gestaltung, die Lernleistung verbessern (vgl. z.B. Groß et al. 2021) was auch für Lernvideos mit KI-Anteilen zu gelten scheint (Jochim et al. 2025). Das enorme Potenzial von KI und aktuellen Large Language Models (LLMs) für die Produktion und Bearbeitung von Videos im Bildungsbereich ist dabei offensichtlich. Der Einsatz von KI verspricht für die oftmals »nebenbei« erfolgende Erstellung von Lernressourcen u.a. Zeitersparnis, technische Qualitätsverbesserung, Inhaltsvertiefung, kreative Gestaltung und nicht zuletzt didaktische Verbesserung (vgl. z.B. de Witt et al., 2020).

Wichtig ist dabei der kritische Umgang mit KI in der Lehre, einschließlich der Reflexion ethischer Fragen und dem Verständnis der Funktionsweise von KI-Systemen. Dazu gehört im Wesentlichen die Kompetenz, KI-generierte Ergebnisse bewerten und deren Plausibilität einschätzen zu können (z.B. Schmohl et al., 2023). Nach Wannemacher et al. (2024) gehören zu den Voraussetzungen für einen erfolgreichen KI-Einsatz in Studium und Lehre ausreichende Zeit- und Personalmittel für Entwicklung, Implementierung und Didaktisierung, eine verlässliche IT-Infrastruktur sowie der Zugang zu datenschutzkonformen Systemen. Ebenso wichtig ist die Qualifizierung der Nutzer:innen, damit Lehrende und Studierende technische Grenzen kennen und die Technologie kompetent und reflektiert einsetzen. Der Umgang mit KI ist dabei eine Basiskompetenz, die erworben werden muss (vgl. Arbeitsgruppe *Digitale Medien und Hochschuldidaktik* der Deutschen Gesellschaft für Hochschuldidaktik 2025). Dementsprechend entstehen zunehmend Angebote in diesem Bereich, wie z.B. der KI-Campus (https://ki-campus.org/).

Hier ordnet sich unser vom BMBFTR gefördertes Projekt *KIViPro* (2024–2026) ein, in dem wir ein modulares Weiterbildungsprogramm zur Produktion von Lehr/Lernvideos in Hochschule und Schule mit Einsatz von KI entwickeln. Es soll in leicht zugänglicher Weise Lehrende an Hochschule und Schule unterstützen, Lehr- und Lernvideos mit Hilfe von KI, seien dies KI-gestützte Software-Tools oder generelle LLMs, effizient zu erstellen, zu bearbeiten und zu veröffentlichen. Das Lernangebot aus dem Projekt *KIViPro* wird als Open Educational Resource (OER) auf TIB TOERN

(Plank et al. 2023) veröffentlicht, ist damit vielseitig nachnutzbar und soll mit der Community vor dem Hintergrund der extrem dynamischen Entwicklung im KI Bereich nachhaltig weiterentwickelt werden. Begleitet wird die Kursentwicklung von virtuellen Evaluations-Workshops.

In diesem Beitrag fokussieren wir auf die Ergebnisse der Bedarfsanalyse und eine Umfrage sowie die Beschreibung des entstehenden Lernangebots. Schließlich diskutieren wir das nicht nur für unser Projekt bedeutsame Spannungsfeld aus extrem dynamischen Inhalten, wie wir sie bei der KI Entwicklung sehen, und statischen Lernangeboten.

Lehrende im Fokus: Umfrageergebnisse

Zur Unterstützung der Bedarfsanalyse zum Zwecke der zielgruppenspezifischen Passgenauigkeit des zu erstellenden Lernangebots wurde im Winter 2024/25 eine Online-Umfrage im Bereich von Hochschule, Wissenschaftsinfrastruktur und Schule durchgeführt. Die mit LimeSurvey erhobenen und standardisierten Fragen (Likert-Skalen, Multiple Choice, offene Textergänzungen) zielten auf bereits vorhandene Erfahrungen und Kompetenzen, die Produktionsziele (Vortrag, Erklärvideo, Interview, Projektvideo etc.) sowie konkrete Wünsche und Anforderungen an das Lernangebot. Es beteiligten sich 104 Personen aus den genannten Bereichen, wobei weit über die Hälfte der Teilnehmer:innen aus dem Bereich Hochschule kamen, während Lehrende aus der Schule stark unterrepräsentiert waren. Zweck und Zielgruppen der Videoproduktion von Hochschullehrenden gehen weitgehend kongruent. »Hochschullehre«, »Vorträge/Präsentationen« und »Unterricht« liegen zwischen 42 % und 50 % bei den Zwecken in etwa gleichauf. Dazu passen die Zielgruppen, bei denen »Studierende« mit 50 % vorn liegen, gefolgt von Wissenschaftler:innen und »Öffentlichkeit/Allgemeinheit«. Bei den Produktionszielen zeigt sich eine große Breite von inhaltlichen und didaktischen Zielen. Dabei dominieren aber die Erklärvideos.

Betrachtet man die bereits vorhandenen Erfahrungen und Kompetenzen der Teilnehmenden, fällt besonders die große Diskrepanz zwischen der insgesamt recht hohen Erfahrung mit herkömmlicher Videoproduktion und der in der Selbsteinschätzung gleichzeitig sehr niedrigen Erfahrung und Kompetenz bei KI gestützter Erstellung von Lehrvideos auf. Der Großteil der Teilnehmenden hat hier keine Erfahrung mit KI in der Videoproduktion. Diese Diskrepanz verdeutlicht pointiert den Bedarf für Lern- und Informationsangebote in diesem Bereich.

Das größte Problem bei der Lehrvideoproduktion ist laut unserer Umfrage der Zeitmangel. Weniger, aber ebenfalls signifikant genannt werden Probleme bei der inhaltlichen Konzeption sowie hinsichtlich der Bild- und Sprachqualität. In diesen Bereichen kann der Einsatz von KI gezielte Unterstützung leisten und das vorge-

stellte Kurskonzept adressiert diese Aspekte explizit. Die offenen Antworten zur Frage nach den Herausforderungen des KI-Einsatzes in der Lehrvideoproduktion zeigen ein breites Spektrum an Themen. Besonders häufig genannt wurden Rechts- und Datenschutzfragen (z.B. »Was passiert mit den Daten, wenn ich sie in die KI einspeise?« oder »Muss ich kenntlich machen, dass KI verwendet wurde?«), gefolgt von technischen und organisatorischen Hürden wie Zugänglichkeit performanter Systeme, Kosten, Toolvielfalt und hohem Zeitaufwand für Einarbeitung und Nachbearbeitung (»Es gibt nicht das Tool, man muss vieles ausprobieren«). Ebenso betont wurden ästhetische und kreative Bedenken, etwa der Verlust von Authentizität, Glaubwürdigkeit und Originalität (»dass es deutlich nach KI aussieht und alles am Ende gleich wirkt«). Die Antworten verdeutlichen insgesamt, dass neben der Klärung grundlegender Fragen auch ein hoher Bedarf an praxisnahen, konkreten Handreichungen besteht, um den Einsatz von KI in der Lehrvideoproduktion reflektiert und kompetent gestalten zu können.

Sowohl die Vielzahl prägnanter und inhaltlich aussagekräftiger Beiträge der Teilnehmenden als auch die im Verhältnis zur begrenzten Stichprobengröße bemerkenswert hohe Beteiligung lieferten wertvolle Einblicke in bestehende Bedarfe und Erwartungen der Community, die als Grundlage für die Ausgestaltung des Lernangebots dienen. Insgesamt wurde der deutliche Bedarf an Lern- und Informationsangeboten zur Erstellung von Lehr- und Lernvideos unter Einsatz von KI, wie in *KIViPro* konzipiert, deutlich sichtbar.

Das Lernangebot

Das Lernangebot zu KI in der Lehrvideoproduktion wird als Online-Kurs auf der Moodle basierten TIB-Lernplattform TOERN bereitgestellt und dauerhaft als *Open Educational Resource* (OER) mit CC-Lizenz verfügbar sein (https://tib.eu/kivipro).

Die inhaltliche Struktur orientiert sich am generellen Produktionsprozess eines (Lern-)Videos sowie etablierten Standards (z.B. Jovy 2021; Langewitz et al. 2024) und gliedert sich in die großen Cluster »Inhalt und Story«, »Planung«, »Produktion«, »Postproduktion«, »Veröffentlichung und Nachnutzung«. Jedes dieser Cluster enthält ca. 4–5 fokussierte Unterseiten zu spezifischen Themen wie »Audio verbessern« oder »Untertitel und Übersetzung«. Diese modularen Seiten sind einzeln auffind- und bearbeitbar, sodass Lernende gezielt die für sie relevanten Inhalte auswählen können, ohne den gesamten Kurs durchlaufen zu müssen. Die einzelnen Module sind bewusst kompakt gehalten, um auch in zeitlich begrenzten Nutzungssituationen schnell zugänglich und bearbeitbar zu sein. Damit adressiert das Angebot auch die in der Umfrage deutlich geäußerte Problematik des Zeitmangels.

Die aufgrund unzureichender personeller und zeitlicher Ressourcen nicht auf einem formalen Modell beruhende didaktische Strukturierung der einzelnen Mo-

dule orientiert sich an Kompetenzen. Jeder Seitenkopf enthält neben einem kurzen inhaltlichen Überblick drei bis fünf Kompetenzformulierungen, die in Anlehnung an die Bloom-Taxonomie (Krathwohl 2002) formuliert sind. Sie ermöglichen eine schnelle Erfassung der Lernziele und eine didaktische Reduzierung auf das Wesentliche. Zusätzlich unterstützen gestalterisch hervorgehobene Elemente wie »Hinweis«, »Stolperfalle« oder »Wusstest Du schon?« die kognitive Aktivierung und inhaltliche Vertiefung, ergänzen den Lernstoff und lockern die Module visuell auf. Inhaltlich stellen die Module dabei generelle Prozesse in den Fokus und nicht primär die Beschreibung des Vorgehens bei einzelnen Tools. Der Kurs bietet keine Software-Gebrauchsanleitungen, sondern will mit der Kompetenz- und Prozessorientierung allgemeines Verständnis sowie konkrete Anwendbarkeit im jeweiligen persönlichen Kontext vermitteln.

Um diese Praxisnähe und die Weiterentwicklung des Kurses zu gewährleisten, setzen wir auf die Einbindung der Community bereits in der Entstehungsphase (u.a. durch Evaluations-Workshops) wie auch nach Veröffentlichung. Dies erfolgt u.a. über Foren und andere kollaborative Strukturen innerhalb des Kurses. Die Communityanbindung dient dabei nicht nur der Qualitätssicherung und Weiterentwicklung, sondern stellt auch einen Lösungsansatz für das grundsätzliche Spannungsfeld zwischen einem extrem dynamischen Thema wie KI und einem klassischen Lernangebot, das im Kern statisch fundiert ist. Wie können wir mit solchen statischen Lernangeboten der derzeitigen Entwicklung von KI noch folgen und gerecht werden? Dieses Problem betrifft letztlich alle aktuellen Lerninhalte und -projekte, die KI als Gegenstand haben (siehe z.B. auch Platte 2025). Neben der genannten Communityeinbindung setzen wir hier zum einen auf regelmäßige Ergänzung und Überarbeitung, die durch die kleinteilige und vor allem prozessorientierte Struktur des Lernangebots möglich wird. Zum anderen etablieren wir eine *Future-Zone*, in der aktuelle Entwicklungen unstrukturiert gesammelt und diskutiert werden können.

Zusammenfassung und Ausblick

KIViPro entwickelt ein modulares Kursangebot zur Produktion von Lehr/Lernvideos in Hochschule und Schule mit Einsatz von KI. Wir haben im Winter 2024/25 104 Lehrende zum Thema KI Einsatz in Schulen und Hochschulen befragt, wobei die Schulen unterrepräsentiert waren. Die Befragung ergab, dass zwar viel klassische Videokompetenzen vorhanden sind, aber es eine geringe KI-Erfahrung bei den Befragten gibt. Beim KI-Einsatz stehen Rechts- und Datenschutzfragen, Kreativität/Ästhetik und Überforderung im Fokus. Gewünscht sind Grundlagenklärungen und konkrete Handreichungen. Darauf aufbauend entwickeln wir *KIViPro* und veröffentlichen es als nachnutzbare OER und mit begleitenden Online Workshops.

Gerade weil KI die Bildung und Wissenslandschaft rasant verändert, behalten aus unserer Sicht statische Lernmodule ihren Wert: Sie schaffen einen gemeinsamen Bezugsrahmen, vermitteln robuste Grundbegriffe, bieten Orientierung in ethischen und rechtlichen Fragen und ermöglichen ein Lernen jenseits von Tool-Hypes. Dieses stabile Fundament ist die Voraussetzung, um Neuerungen sinnvoll einzuordnen und souverän anzuwenden. Gleichzeitig wird unser Format nicht starr bleiben, denn auf der Basis des stabilen Fundaments im Hinblick auf Lernziele und zentrale Kompetenzen wollen wir Beispiele, Werkzeuge und Fallstudien regelmäßig erneuern und so dynamisch halten. Zu diesem Zweck bauen wir auf eine modulare Architektur, eine kontinuierliche Community-Einbindung sowie spezielle dynamische Bereiche wie die *Future-Zone*. Darüber hinaus werden perspektivisch geplant und bereits experimentell erprobt die Integration eines KI-Chatbots, der in den einzelnen Modulen Inhalte zusammenfassen und ergänzen oder bei Bedarf weitere Informationen geben kann, sowie die Integration von KI-gestützten Videotools, mit denen live im Kurs der textuelle Inhalt praktisch erprobt und visualisiert werden könnte.

Literatur

Arbeitsgruppe *Digitale Medien und Hochschuldidaktik* der Deutschen Gesellschaft für Hochschuldidaktik (2024): Didaktische Handreichung zur praktischen Nutzung von KI in der Lehre. https://www.gmw-online.de/wp-content/uploads/2024/10/KI-Handreichung-dghd_GMW_V01_21102024.pdf.

de Witt, Claudia/Rampelt, Florian/Pinkwart, Niels (2020): Künstliche Intelligenz in der Hochschulbildung. Whitepaper https://doi.org/10.5281/zenodo.4063722.

Groß, Nele/Preiß, Jennifer/Paul, Daria/Brase, Alexa/Reinmann, Gabi (Hg.) (2022): Student Crowd Research. Videobasiertes Lernen durch Forschung zur Nachhaltigkeit, Münster/New York: Waxmann. https://doi.org/10.25656/01:25824.

Jochim, Julia/Post, Cecilia/Bolten-Bühler, Ricarda (2025): »Unterrichtet vom Avatar? Studierenden-Präferenzen bei Erklärvideos«, in: MedienPädagogik 65 (MEDIDA24), S. 69–93.

Jovy, Jörg (2021): Digital filmen: Das umfassende Handbuch (4., aktualisierte Auflage), Bonn: Rheinwerk Verlag.

Krathwohl, David R. (2002): »A Revision of Bloom's Taxonomy: An Overview«, in: Theory Into Practice, 41(4), S. 212–218.

Langewitz, Oliver/Schlegel, Kurt/Schlote, Elke (2024): Wirkungsvolle Videos fürs Web für Dummies, Weinheim: Wiley-VCH.

Plank, Margret/Klinger, Axel/Beutnagel, Britta (2023): The Project TOERN – Unlocking the Potential of Open Educational Resources on the German »(digital)

networked infrastructure for education«, in: e-learning and education, Iss. 15. https://nbn-resolving.org/urn:nbn:de:0009-5-57778.

Platte, Laura (2025):/imagine a Game about AI – Entwicklung des Serious Games »Singularity« über und mit generativer KI. Vortrag auf dem University: Future Festival 2025. https://www.youtube.com/watch?v=zi5zEDfWhuU&t=396s vom 23.05.2025.

Schmohl, Tobias/Watanabe, Alice/Schelling, Kathrin (Hg.) (2023): Künstliche Intelligenz in der Hochschulbildung. Chancen und Grenzen des KI-gestützten Lernens und Lehrens, Bielefeld: transcript Verlag.

von Garrel, Joerg/Mayer, Jana (2025): Künstliche Intelligenz im Studium. Eine quantitative Längsschnittstudie zur Nutzung KI-basierter Tools durch Studierende (2023 & 2025), Darmstadt: Hochschule Darmstadt.

Wannemacher, Klaus/Bosse, Elke/Lübcke, Maren/Kaemena, Alena (2025): Wie KI-Studium und Lehre verändert. Anwendungsfelder, Use-Cases und Gelingensbedingungen. Arbeitspapier 87, Berlin: Hochschulforum Digitalisierung.

Singularity – Ein Serious Game für AI Literacy[1]

Laura Platte und Malte Persike[2]

*Im Projekt Singularity – A Journey to AI Literacy entwickelt das Center für Lehr- und Lernservices (CLS) der RWTH Aachen University ein Serious Game, das Studierende aller Fächer spielbasiert an die Grundlagen der AI Literacy heranführt. Die Lernenden finden sich in der Rolle als Notfall-Operator*in auf einem Raumschiff und interagieren mit verschiedenen historischen und modernen KI-Systemen. Ziel ist es, Kompetenzen im kritischen Umgang mit KI, im Anwenden von Prompting-Techniken und im Hinterfragen KI-generierter Informationen zu fördern. Das Spiel integriert unterschiedliche didaktische Formate wie Videos, Texte, Quizzes, interaktive Aufgaben und Storytelling. Bei der Spieleentwicklung experimentieren wir als Projektteam bewusst mit dem Einsatz generativer KI, etwa in Form von bildgenerativen Modellen bei der Entwicklung der Spielgrafik oder durch die Einbindung eines textgenerativen Modells als Teil des Spiels.*

Singularity – A Serious Game for AI Literacy

In the project Singularity – A Journey to AI Literacy, the Center for Teaching and Learning Services (CLS) at RWTH Aachen University is developing a serious game that playfully introduces students from all disciplines to the basics of AI literacy. Players take on the role of an emergency operator aboard a spaceship and interact with various historical and modern AI systems. The game aims to foster competencies in critically engaging with AI, applying prompting techniques, and questioning AI-generated information. It integrates diverse didactic formats such as videos, texts, quizzes, interactive tasks, and storytelling. During development, the project team deliberately experiments with the use of generative AI – for instance, by employing image-generating models for the game's graphics and incorporating a text-generative model as part of the gameplay experience.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000–0002-7825-089X

Setting und Story des Spiels

Das browser-basierte Serious Game Singularity – A Journey to AI Literacy soll Studierende aller Fächer spielerisch an die Grundlagen der AI Literacy heranführen. Zu Beginn des Spiels finden sich die Lernenden als Notfall-Operator*in an Bord eines ansonsten menschenleeren Raumschiffs wieder. Die Mission besteht darin, eine Gruppe von genetisch weiterentwickelten Waschbären auf einem fernen Planeten auszusetzen. Die Lernenden wurden aus dem Kälteschlaf geweckt, um verschiedene unerwartete Probleme zu beheben, die durch die Waschbären, aber auch durch die unausgereiften technischen Lösungen zur Automatisierung des Raumschiffs entstehen. Als Begleiterin und Mentorin steht den Lernenden die sprechende Hündin Velka zur Seite (siehe Abbildung 1).

Abbildung 1: Maschinenraum des Raumschiffs (Grafiken generiert mit Midjourney. Nachbearbeitung und Szenendesign: Frida Klein).

Gemeinsam mit Velka erkunden Spielende ihre Umgebung Raum für Raum und begegnen dabei verschiedenen Systemen künstlicher Intelligenz (KI): So begleitet der historische Chatbot ELIZA (vgl. Weizenbaum 1966) das Aufwachen aus dem Kälteschlaf; in einem anderen Raum spielt der Schachcomputer Deep Blue (vgl. IBM o.J.) eine Partie Schach. Im weiteren Spielverlauf üben die Lernenden Prompting-Techniken für textgenerative Modelle ein, erleben die möglichen negativen Auswirkungen der unzureichenden Faktentreue von Large Language Modellen und führen letztendlich die Raummission zum Erfolg.

Was verstehen wir unter AI Literacy?

AI Literacy wurde bereits vor der rasanten Entwicklung und Verbreitung generativer Modelle als zunehmend wichtige überfachliche Kompetenz diskutiert. Long und Magerko (2020) definieren AI Literacy als »Set an Kompetenzen, das Individuen dazu befähigt, KI-Technologien kritisch zu bewerten; effektiv mit KI zusammenzuarbeiten und zu kommunizieren und KI online, zu Hause und am Arbeitsplatz als Werkzeug zu benutzen« (Long/Magerko 2020: 2; eigene Übersetzung). In den letzten Jahren haben Publikationen zu diesem Thema deutlich zugenommen und es besteht Einigkeit darüber, dass diese Kompetenz in der Hochschulbildung adressiert werden muss (vgl. Laupichler et al. 2022). Ein einheitliches Verständnis von AI Literacy oder gar ein etablierter Kompetenzrahmen existiert jedoch noch nicht. Verschiedene Ansätze unterschieden sich etwa in der Frage, welche Rolle Programmierfähigkeiten und Ethik spielen sollten, wie Kenntnisse und Anwendung gewichtet werden und inwiefern es sich tatsächlich um eine eigene Kompetenz oder lediglich um eine neue Wissensdomäne handelt. Zudem legen die Ansätze einen unterschiedlich starken Fokus auf generative Technologien.

Angesichts der beschriebenen Ambiguitäten bei Kompetenzbegriffen zur AI Literacy verorten wir unser Vorhaben wie folgt:

Breites Verständnis von künstlicher Intelligenz: Wir beschränken uns nicht auf generative Technologien, sondern berücksichtigen auf Ebene der Kenntnisse auch die historische Einordnung verschiedener Methoden und Anwendungen sowie vereinfachte allgemeine Modelle von Informationsverarbeitung durch KI. Die Lernenden sollen in die Lage versetzt werden, die aktuellen Entwicklungen im Bereich generativer KI-Anwendungen sowie den Diskurs über ihre Potentiale und Risiken zu kontextualisieren.

Strategien zur Anwendung textgenerativer Modelle: Lernende sollen grundlegende Prompting-Strategien (vgl. Schulhoff et al. 2025) zur effektiven Arbeit mit generativen Technologien einüben und verschiedene Beispiele für mögliche Einsatzszenarien erproben.

Kritisches Hinterfragen von Information: Lernende sollen Informationen, die sie aus (generativen) KI-Anwendungen erhalten, hinterfragen und überprüfen (vgl. Reinmann 2023). Darüber hinaus wird an verschiedenen Stellen im Spiel die einseitige, überoptimistische Darstellung der Potentiale von KI-Technologien aufgezeigt und mit historischen Fehleinschätzungen sowie Fehlfunktionen von Systemen auf dem Raumschiff kontrastiert.

Eine Besonderheit am Konstrukt AI Literacy ist zudem, dass sich einige konkrete Lernergebnisse im Bereich der Kenntnisse über generative Modelle oder auch der Prompting-Techniken im Zuge der rapiden technischen Entwicklung vergleichsweise schnell verändern. In unserer Projektarbeit bewegen wir uns daher bei einigen Lernzielen in einer feinen Abwägung zwischen momentaner Relevanz

und einer ›Halbwertszeit‹ der Inhalte. Unsere wichtigste Strategie ist hierbei die Modularisierung des Spiels: Das Raumschiff besteht aus verschiedenen Räumen, die jeweils weitgehend abgeschlossene Spielszenen und Lernressourcen beinhalten. So können Spielinhalte mit vergleichsweise geringem Aufwand ausgetauscht oder erweitert werden.

Warum ein Serious Game?

Ein Serious Game ist ein Spiel, das über den reinen Unterhaltungswert hinausgeht und explizite pädagogische Ziele verfolgt (vgl. Abt 1970). Es kombiniert also spielerische Elemente mit »ernsthaften«, lernbezogenen Inhalten und Erfahrungen.

Kapp (2016) arbeitet drei Elemente heraus, die charakteristisch für Spiele sind: Engagement, Autonomy und Progression. Diese spielerischen Elemente müssen in einem Serious Game effektiv mit den lernbezogenen integriert sein statt nur nebeneinander zu stehen (vgl. Caserman et al. 2020). Dies lässt sich anhand des Spieldesigns von Singularity illustrieren:

Engagement bezeichnet die aktive Einbindung von Lernenden, z.B. durch Storytelling oder Rätsel. In Singularity arbeiten wir stark mit der Story und nutzen sie nicht nur, um einen narrativen Rahmen zu schaffen: Die diversen Probleme mit KI-Technologien, mit denen die Lernenden auf dem Raumschiff umgehen müssen, werden als Beispiel für einen übertreibenden Technikoptimismus aufgegriffen und in den Spieldialogen kritisch eingeordnet.

Autonomy meint die Möglichkeit, im Spiel innerhalb bestimmter Regeln selbstbestimmte Entscheidungen treffen zu können und Raum zum Ausprobieren und sogar zum Scheitern zu haben. In Singularity wechseln sich vorgegebene Spielsequenzen mit solchen ab, in denen die Lernenden durch freie Exploration eigene Erfahrungen machen. So erleben sie direkt am Anfang des Spiels den historischen Chatbot ELIZA, mit dem sie durch freie Texteingabe interagieren. Tatsächlich stellt sich ELIZA in der betreffenden Spielsituation als nutzlos heraus. Die Lernenden machen also die Erfahrung, dass nicht jede Situation durch den Einsatz verfügbarer Technologien zu lösen ist.

Progression bezieht sich auf richtungsweisende Strukturen wie Spielziel(e), Levels, Belohnungen und Feedback. Der Spielfortschritt in Singularity wird durch verschiedene narrative und auch visuelle Feedbacks signalisiert. Ein zentraler Mechanismus ist das Öffnen weiterer Räume des Raumschiffs und das Sammeln von KI-Modulen, auf die die Lernenden zurückgreifen können.

Trotz langjähriger Forschung zur Wirksamkeit von Games gibt es in einigen Fragen noch keine eindeutigen Ergebnisse. In Bezug auf die kognitiven Wirkungen können die meisten untersuchten Spiele mindestens mit anderen digitalen Lernformaten mithalten. In einigen Fällen erzielen sie sogar bessere Ergebnisse, wobei die

Effektstärken im kleinen bis moderaten Bereich liegen (vgl. Gentry et al. 2019; Lamb et al. 2018). Besondere Potentiale zeigen Serious Games in vielen Studien in der affektiven Wirkung, etwa in der Motivation (vgl. Arias-Calderón et al. 2022) und im Selbstwirksamkeitserleben (vgl. Perlwitz/Stemmann 2022) der Lernenden. Erwähnt sei jedoch, dass oft schwer unterscheidbar ist, inwiefern Effekte auch durch inhaltsunabhängige Faktoren bedingt sind. Diskutiert werden in diesem Zusammenhang etwa eine hohe Verweildauer in Serious Games und damit eine im Vergleich zu traditionellen Lernmedien signifikant erhöhte Lernzeit (vgl. Hummel et al. 2024) sowie der Novelty-Effekt, also eine zunächst erhöhte Motivation durch den Neuheitswert des Formats (vgl. Romero Rodríguez et al. 2025). Der aktuelle Stand der Forschung zeigt also Potentiale auf, mahnt aber gleichzeitig zu einem sorgfältigen Spieldesign und einer begleitenden Evaluation kritischer Faktoren.

Für Singularity sind insbesondere die Potentiale durch das freie Explorieren (im Sinne der Autonomy) und die positive affektive Wirkung von Bedeutung. Da sich viele Studierende aktuell selbstständig und außercurricular mit AI Literacy beschäftigen (vgl. Budde/Friedrich 2024), kann ein Serious Game einen motivierenden, niederschwelligen Einstieg bieten.

Didaktische Formate

Das Spiel integriert verschiedene didaktische Formate: Neben klassischen Informationsmedien wie kurzen Lernvideos, Texten und erklärenden Abbildungen dienen an einigen Stellen auch die Story und Dialoge dazu, Erkenntnisse zu entwickeln und hervorzuheben. Quizzes und interaktive Drag-and-Drop-Abbildungen sollen Wissen durch Wiederholung und aktive Interaktion verfestigen. Quizformate können außerdem – im Sinne der Autonomy – als Raum zum spielerischen Scheitern dienen: In einem Quiz sollen Lernende identifizieren, welches von drei Bildern künstlich generiert wurde, was an dieser Stelle nicht möglich ist. Durch diese Erfahrung und die Erläuterungen im Spieldialog sollen die Lernenden verinnerlichen, wie wichtig es ist, KI-generierte Bilder gewissenhaft als solche zu kennzeichnen.

Für das Erproben von Prompting-Techniken arbeiten wir außerdem an der Einbindung eines generativen Modells.

Spielentwicklung mit generativen Technologien

Im Prozess der Spielentwicklung experimentieren wir bewusst damit, text- und bildgenerative Technologien nicht nur als Thema des Spiels zu fokussieren, sondern auch an verschiedenen Stellen der Spielproduktion einzubinden. Mithilfe von textgenerativen Modellen wie dem OpenAI-basierten RWTHgpt haben wir

beispielsweise Ideen zur Spielstory gebrainstormt oder Texte für Formate wie Videos oder Quizzes optimiert. Auch in der Spielgrafik – also Hintergrundgrafiken, Charaktere und Objekte – arbeiten wir umfangreich mit der Plattform Midjourney. Andere Arbeitsschritte sind jedoch weiterhin rein »menschlich«, etwa das Zeichnen der Storyboards in der Entwicklung des Szenendesigns (siehe Abb. 2).

Abbildung 2: Ausschnitt aus einem Storyboard (Grafik: Frida Klein).

Alle Ideen, Texte und Grafiken werden in regelmäßigen Terminen im Team gereviewt und angepasst. Dabei liegt das Augenmerkt zum einen auf der Konsistenz: So werden beispielsweise alle Hintergrundgrafiken händisch nachbearbeitet, um visuelle Details wie die Raumschiffarchitektur zu vereinheitlichen oder ablenkende Bildelemente zu reduzieren. Zum anderen müssen lerngegenstandsbezogene Kriterien berücksichtigt und im Zweifel gegenüber den spielbezogenen Überlegungen priorisiert werden (vgl. Caserman et al., 2020). Ein Beispiel hierfür ist der Companion-Charakter, der sowohl Hilfe in der Spielnavigation anbietet als auch Lerninhalte erklärt. Sowohl im Brainstorming im Team als auch mit RWTHgpt entstand wiederholt der Vorschlag eines niedlichen Roboters im Stil von R2D2 in Star Wars. Dies stünde jedoch in Konflikt zum angestrebten Lernergebnis, Informationen aus technischen Systemen – sei es Roboter oder GPT-Modell – immer kritisch zu überprüfen. Die Lernenden werden daher nun von der sprechenden Hündin Velka begleitet, die in Abbildung 1 zu sehen ist.

Fazit und Ausblick

Unsere bisherigen Erfahrungen in der Spielentwicklung zeigen, dass Serious Games einen vielversprechenden Ansatz darstellen, die bestehenden Bildungsangebote in diesem Bereich zu diversifizieren und Lernende einzuladen, die Grundlagen von AI Literacy spielerisch zu erkunden. Der gezielte Einsatz generativer KI kann dabei den anspruchsvollen, zeit- und ressourcenintensiven Entwicklungsprozess eines Spiels wirksam unterstützen.

Wie im Abschnitt zu Serious Games erläutert, stellt die Entwicklung eines lernwirksamen Spiels hohe Anforderungen an das Spieldesign sowie an die Konzeption

und Integration didaktischer Formate. Im Rahmen einer Evaluation mit Studierenden sollen daher unter anderem Verständlichkeit, Lernzeit, Spielerlebnis und Lerneffekte untersucht werden, bevor das Spiel einer größeren Zielgruppe, etwa über den KI-Campus, zugänglich gemacht werden soll.

Danksagung und Förderhinweis

Als Autor*innen stehen wir stellvertretend für ein großes Team an Beteiligten. Besonders bedanken möchten wir uns bei Tobias Grimm und Lisa Bachmann (Serious Game-Entwicklung) sowie Frida Klein (Grafikdesign).

Singularity – A Journey to AI Literacy wird im Rahmen eines gleichnamigen zweijährigen Projekts am Center für Lehr- und Lernservices der RWTH Aachen entwickelt. Die Finanzierung erfolgt über die *Freiraum2023*-Förderlinie der Stiftung Innovation in der Hochschullehre.

Literatur

Abt, Clark C. (1970): Serious games, New York: Viking Press.

Arias-Calderón, Manuel/Castro, Javiera/Gayol, Silvina (2022): »Serious Games as a Method for Enhancing Learning Engagement: Student Perception on Online Higher Education During COVID-19«, in: Frontiers in Psychology, 13, S. 1–11.

Budde, Jannica/Friedrich, Julius-David (2024): Monitor Digitalisierung 360° Wo stehen die deutschen Hochschulen? Arbeitspapier Nr. 83. Berlin: Hochschulforum Digitalisierung.

Caserman, Polona/Hoffmann, Katrin/Müller, Philipp/Schaub, Marcel/Straßburg, Katharina/Wiemeyer, Josef/Bruder, Regina/Göbel, Stefan (2020): »Quality Criteria for Serious Games: Serious Part, Game Part, and Balance«, in: JMIR Serious Games, 8(3), S. 1–14.

Gentry, Sarah V./Gauthier, Andrea/L'Estrade Ehrstrom, Beatrice/Wortley, David/Lilienthal, Anneliese/Tudor Car, Lorainne/Dauwels-Okutsu, Shoko/Nikolaou, Charoula K./Zary, Nabil/Campbell, James/Car, Josip (2019): »Serious Gaming and Gamification Education in Health Professions: Systematic Review«, in: Journal of Medical Internet Research, 21(3), S. 1–20.

Hummel, Hans G. K./Nadolski, Rob/Huurdeman, Hugi/Van Lankveld, Giel/Georgiadis, Konstantinos, Slootmaker, Aad/Kurvers, Hub/Hummel, Mick/Neessen, Petra/Van Den Boomen, Johan/Pat-El, Ron/Fischmann, Julia (2024): »Gamebrics: Design, implementation and practical evaluation of analytical rubrics in serious game play«, in: Journal of Computer Assisted Learning, 40(5), S. 2150–2165.

IBM (o.J.): Deep Blue [Unternehmenswebsite]. IBM. https://www.ibm.com/history/deep-blue

Kapp, Karl M. (2016): »Gamification designs for instruction«, in: Charles. M. Rigeluth,/Brina J. Beatty/Rodney D. Myers (Hg.): Instructional-Design Theories and Models: Bd. IV: The Learner-Centered Paradigm Of Education, New York: Routledge, S. 351–384.

Lamb, Richard L./Annetta, Leonard/Firestone, Jonah/Etopio, Elisabeth (2018): »A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations«, in: Computers in Human Behavior, 80, S. 158–167.

Laupichler, Matthias C./Aster, Alexandra/Schirch, Jana/Raupach, Tobias (2022): »Artificial intelligence literacy in higher and adult education: A scoping literature review«, in: Computers and Education: Artificial Intelligence, 3, S. 1–15.

Long, Duri/Magerko, Brian (2020): »What is AI Literacy? Competencies and Design Considerations«, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, S. 1–16.

Perlwitz, Phoebe/Stemmann, Jennifer (2022): »Flow and self-efficacy in a serious game for STEM education«, in: Joint International Conference on Serious Games, Cham: Springer International Publishing, S. 3–16.

Reinmann, Gabi. (2023): Deskilling durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik. Diskussionspapier Nr. 25, Berlin: Hochschulforum Digitalisierung.

Romero Rodríguez, Laura/Sánchez-Alzola, Alberto/Salazar, Alejandro (2025): »Serious Games in Engineering Education: Assessing Novelty Effects and the Influence of Prior Gaming Experience«, in: Computer Applications in Engineering Education, 33(3), S. 1–23.

Schulhoff, Sander/Ilie, Michael/Balepur, Nishant/Kahadze, Konstantine/Liu, Amanda/Si, Chenglei/Li, Yinheng/Gupta, Aayush/Han, HyoJung/ Schulhoff, Svien/Sandeep, Pranav Dulepet/Vidyadhara, Saurav/Ki, Dayeon/Agrawal, Sweta/Pham, Chau/Kroiz, Gerson/Li, Feileen/Tao Hudson/Srivastava, Ashay/Da Costa, Hevander/Gupta, Saloni/Rogers, Megan L./Goncearenco, Inna/Sarli, Giuseppe/Galynker, Igor/Peskoff, Denis/Carpuat, Marine/White, Jules/Anadkat, Shyamal/Hoyle, Alexander/Resnik, Philip (2025): The Prompt Report: A Systematic Survey of Prompt Engineering Techniques, No. arXiv:2406.06608, arXiv.

Weizenbaum, Josef (1966): »ELIZA—a computer program for the study of natural language communication between man and machine«, in: Communications of the AC, 9(1), S. 36–45.

Aus der Rolle gefallen?

Entwicklung von KI-gesteuerten Spielfiguren mit Studierenden[1]

Luke West[2] und Esther Fink[3]

Die an der Universität Tübingen entwickelte Spieleplattform Graveler ermöglicht Lehrenden und Studierenden ohne Programmierkenntnisse Spiele und Rollenspiele für Lehrzwecke zu entwickeln. Mit KI-gesteuerten Spielfiguren (KI-NPCs) können Studierende alltägliche Situationen, Probleme oder Konflikte in einer sicheren Umgebung durchspielen und analysieren, ohne Druck sich zu exponieren oder beurteilt zu werden. Allerdings sind KI-gesteuerte Rollenporträts nicht unproblematisch und können leicht durch unpräzise Prompts oder Bias beeinflusst werden. Gerade das macht die Entwicklung von KI-NPCs aber zu einer spannenden Lernerfahrung für Studierende, die sie herausfordert, ihr Verständnis von Rollen und Szenarien zu reflektieren und zu diskutieren. Wir berichten über ein Lernszenario, das dazu anregt, Rollenspielszenarien mit generativer KI zu entwickeln und Ergebnisse auf Effektivität und ethische Überlegungen hin zu bewerten. Im Bereich Global Awareness Education entwickeln Studierende KI-NPCs im Kontext der Migration nach dem Zweiten Weltkrieg, als Spiegel für die heutige Migrationsdebatte.

Breaking character? – Development of AI-Powered Game Characters with Students

The Graveler game platform developed at the University of Tübingen, enables teachers and students to create games and role-playing games for teaching purposes, even without programming skills. With AI-controlled game characters (AI-NPCs), students can simulate and analyse everyday situations, problems or conflicts in a safe environment without the pressure of exposing themselves or being judged. However, AI-driven character portraits are not without challenges and are easily affected by ambiguous prompts or bias. This is precisely what makes the development of AI-NPCs an exciting learning experience for students, challenging them to reflect on and discuss their understanding of roles and scenarios. We report on a learning scenario from Global Awareness Education, that encourages the development of role-play scripts with generative AI, evaluate its effectiveness and discuss ethical considerations. In this

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000–0003-0559-7865

3 ORCID: 0009–0005-0995-8340

learning activity students develop AI NPCs in the context of migration after the Second World War to reflect on today's migration debate.

Mehr als Unterhaltung: Die kulturelle Bedeutung digitaler Spiele

Spiele sind seit alters her eine kulturelle Praxis, die soziale Konventionen und Regeln vermittelt und die Auseinandersetzung mit Gegenwart und Vergangenheit fördert. Digitale Spiele machen in virtuellen Welten Geschichte und Wissen erlebbar und ermöglichen es Spielenden, durch Zeit und Raum zu reisen, um reale und fiktive Welten aus verschiedenen Perspektiven zu entdecken. Sie sind seit zwei Generationen für viele ein selbstverständlicher Teil des Lebens. Darüber hinaus haben sie sich zu einer eigenständigen kreativen Ausdrucksform entwickelt (Clark 2024).

GBL & Serious Games

Game-based Learning hat das Potential Interesse zu fesseln und die Motivation der Studierenden zu fördern. Serious Games regen zu aktivem und forschendem Lernen an, und verbessern die Lernergebnisse sowohl kurz- als auch langfristig (Wouters et al. 2013), insbesondere dann, wenn die Spiele mit realen Situationen und Herausforderungen verknüpft sind. Spielbasiertes Lernen fördert Fähigkeiten wie kritisches Denken und zwischenmenschliche Kommunikation (Lombardi/Oblinger 2007). Digitale Serious Games machen Szenarien und Situationen erlebbar, die aufgrund von Zugangs- oder Sicherheitsrestriktionen schwer zugänglich sind, und bieten vielfältige multisensorische Erfahrungen (Whitton, 2014). In Deutschland haben Serious Games bereits einen festen Platz in der praxisorientierten Lehrerausbildung (Rüth/Kaspar 2024) sowie in der Hochschul- und Weiterbildung erobert (Deckert et al. 2018) – insbesondere in den Bereichen der Medizin (Bykova/Frank 2016) und Ingenieurswissenschaften (Söbke/Reichelt 2018). Spielerische Erkundung ist ein starker Katalysator für Kreativität und neue Entdeckungen – sei es auf persönlicher, beruflicher oder akademischer Ebene (James/Nerantzi 2019). Noch wirkungsvoller im Hinblick auf Motivation und Schlüsselqualifikationen sind Lernerfahrungen, bei denen die Studierenden selbst gestalten und eigene Spiele entwickeln, anstatt nur ein Spiel zu spielen (Whitton 2012).

AI NPCs & Immersive Learning

Non-Player-Characters (NPCs) sind skriptgesteuerte Spielfiguren, die die Spielenden in die Erzählung eines Spiels einbinden, indem sie Aufgaben stellen, Infor-

mationen und Hinweise liefern und auf die Aktionen der Spielenden reagieren. Die wissenschaftliche Auseinandersetzung mit adaptiven KI-gesteuerten NPCs in virtuellen Rollenspielen ist noch lückenhaft, solange sich KI-NPCs aber glaubwürdig und rollenadäquat verhalten, werden sie als eigenständig und intelligent wahrgenommen (da Silva/da Souza 2021). KI-gestützte NPCs können Spielenden vielfältig Interaktionsmöglichkeiten bieten und ermöglichen emotionale Erfahrungen, die zuvor nur im realen Raum möglich waren (Schwing 2025). Mit Hilfe von KI-NPCs können alltägliche Situationen, Probleme oder Konflikte in einer sicheren Umgebung durchgespielt und analysiert werden, ohne dass Studierende durch Rollenspiele vor Publikum gehemmt sind (Gonzales et al. 2024). Allerdings ist dies nicht unproblematisch: durch bestehende Sprachmuster oder gesellschaftliche Stereotypen kann es zu Voreingenommenheit in KI-Algorithmen oder Trainingsdatensätzen kommen. KI-gesteuerte Figurenporträts können ungenau sein, von der geplanten Handlung abweichen, und Nuancen der Figuren gehen leicht verloren. Figuren können aus der Rolle fallen oder unlogisch handeln, weil sie nicht genügend Informationen über die Rolle haben oder weil sie auf Wissen außerhalb des Rollenspielszenarios zurückgreifen. In realen Szenarien geht es oft um komplexe Probleme und unerwartete Situationen, die sich nicht ohne Weiteres nachstellen lassen und auf Hintergrundinformationen beruhen, die leicht fehlinterpretiert werden können.

Dieses Verzerrungspotential bei KI-gestützten Rollenspielen macht daher Lernaktivitäten, die über den simulierten Dialog hinausgehen, wie etwa das gemeinsame Erarbeiten neuer Rollen und Szenarien oder die Reflexion der gemachten Erfahrungen besonders wichtig. Sie sind es, die die kritische KI-Kompetenz fördern (Dillig et al. 2024). Yuan et al. (2025) sehen großes Potenzial in KI-gestütztem immersivem Rollenspiel mit historischen Figuren, deren Dialoge auf der Grundlage historisch akkurater Datenbanken generiert werden, um eine möglichst authentische Darstellung zu gewährleisten. Steinmaurer et al. (2024) entwickelten ein Modell für immersiven Geschichtsunterricht mit LLMs (iHELLM), das ebenfalls historische Quellen und Kontexte integriert, um die immersive Lernumgebung zu verbessern. Gleichzeitig birgt der Einsatz von KI bei der Darstellung historischer Figuren die Gefahr von Fehlinformation oder unethischer Darstellung, durch Bias, Filter oder algorithmische Verzerrung, insbesondere für Angehörige und Opfergemeinschaften deren Geschichte sensibel oder traumatisch ist. Wie sich KI-gestützte Dialoge entwickeln, lässt sich möglicherweise nur schwer vorhersagen, oder Informationen können auf dem Weg von Fachleuten zu Softwareentwicklern verloren gehen oder falsch dargestellt werden (Shorey et al. 2019). Andererseits zeigt das Feedback der Studierenden, dass die spielerische Form die Lernerfahrung besonders interessant macht, und Studierende zudem kritisch denken, Probleme lösen und ihr Verständnis von Rollen und Szenarien diskutieren müssen. Vor dem Hintergrund von Holocaust Zeitzeugenaussagen empfehlen Kozlovski & Makhortykh, (2025) daher

KI-gestützte digitale Figuren auf der Basis folgender Kriterien zu bewerten authentische Präsenz, Einverständnis der Betroffenen, positiver Nutzen, Transparenz und Minderung von Schadensrisiko.

Gamedesign als didaktisches Prinzip

Serious Games bieten nicht nur Lernpotenzial, sondern auch großes Gestaltungspotenzial (Marone 2016). Die Entwicklung solcher Spiele fördert neben strukturierten Lernzielen auch selbstgesteuerte, erfahrungsbasierte Lernprozesse und persönliche, soziale und kognitive Kompetenzen (Kafai/Burke 2015), und macht Game Design zu einem vielversprechenden Lehrformat in der Hochschulbildung. Es verbindet formelles und informelles Lernen und fördert über ko-kreative Designprozesse zugleich praxisrelevante Haltungen, Kompetenzen und Ausdrucksformen wie Empathie, Sinn- und Bedeutungsorientierung, Kreativität und Teamarbeit für die berufliche Zukunft (Arnab et al. 2019; Yang et al. 2024).

Die Software Graveler (Körner et al. 2024) wurde als Lernplattform konzipiert, die die Erstellung von Spielen und Rollenspielen für Lehrzwecke ermöglicht. Nutzende können einfach interaktive webbasierte virtuelle 2D-Welten erstellen, die die Spielenden dazu herausfordern, Welten zu erkunden, Aufgaben zu erfüllen und Herausforderungen zu meistern. Die Spielenden können mit der Spielwelt, anderen Spielenden (Avataren) und KI-gesteuerten Non-Player-Charakteren (KI-NPCs) auf vielfältige Weise interagieren, sowohl in fiktionalen als auch in realen Szenarien. Graveler unterstützt artefaktbasierte Interaktionen, Diskurse und soziale Interaktionen und ermöglicht damit Studierenden und Lehrenden, auf einfache Weise interaktive Lernerfahrungen zu erstellen, zu teilen, wiederzuverwenden und anzupassen. Die Plattform wurde zunächst für die Lehrveranstaltung »What is Globalization« des Bereichs Global Awareness Education an der Universität Tübingen konzipiert, die sich mit den vielfältigen Auswirkungen der Globalisierung auf Politik, Kulturen, Umwelt und zwischenmenschliche Interaktionen beschäftigt. Darüber hinaus wurden bereits eine Reihe weiterer Spiele wie »Geschichte der antiken und chinesischen Numismatik« und »Klimawandel« entwickelt und umgesetzt. Angesichts der immersiven Lernvorteile und ethischen Bedenken KI-basierter Rollenspiele in Serious Games sowie der Lernvorteile der gemeinsamen Spieleentwicklung bedarf es eines kreativen Raums, in dem Studierende aktiv an Gestaltung, Reflexion und Diskussion teilnehmen können. Vor diesem Hintergrund dient Graveler als Spieleentwicklungsplattform, auf der KI-basierte Rollenspiele nicht nur gespielt, sondern auch kollaborativ entwickelt werden können. Die Artefakte bleiben im Lernkontext erhalten und die Studierenden haben die Möglichkeit, ihre promptgesteuerten KI-NPCs zu testen, deren Rollentreue zu reflektieren und zu überarbeiten.

Mit KI-gesteuerten NPCs in die Geschichte eintauchen

Hier berichten wir über ein Lernszenario in dem Studierende mit Graveler KI-NPCs für »Displaced Persons« im zweiten Weltkrieg entwickeln, um dann die Charaktere und deren Dialoge kritisch auf ihre Wirkung und ethische Aspekte zu prüfen. Dabei reflektieren sie die Qualität ihrer Prompts und den verantwortungsvollen KI-Einsatz. Da das Seminar online stattfand, wurde die Aktivität synchron über Videokonferenz durchgeführt und die Lernenden teilten ihre Hintergrundrecherche, KI-NPC-Beschreibungen, Bewertungen und Überlegungen über ein virtuelles Whiteboard mit Lehrenden und anderen Studierenden. Das Graveler NPC-Dashboard ermöglichte Studierenden ein Spielszenario zu beschreiben und Mechaniken und Interaktionen für den KI-NPC festzulegen und den Charakter anschließend im Dialog mit verschiedenen LLM-Modellen zu testen. Die Prompt-Struktur bestand aus vier Teilen: (1) dem Shell-Prompt, der festlegt, dass der Agent immer mit 1 bis 2 kurzen Sätzen antwortet, (2) der Szenariobeschreibung einschließlich historischer Referenzen, fiktiver Ausarbeitung und Verbindung zum Spielkontext, (3) der vom Spielenden während des Rollenspiels eingegebenen Dialoge und (4) den letzten 3 Dialogrunden als Konversationsgedächtnis des Agenten.

Lehren und Lernen durch Game Design im Hochschulalltag

Elf Studierende nahmen an der KI-NPC-Designaktivität im Rahmen eines englischsprachigen Blockkurses zum Thema Gamedesign teil. Zunächst erhielten die Studierenden eine Einführung in das Serious Game »PostWar«, das gemeinsam von Studierenden und Lehrenden an der Universität Tübingen entwickelt wurde. Die Geschichte spielt im frühen Nachkriegsdeutschland, wo ein Postangestellter Vertriebene durch die Zustellung von Briefen und Gegenständen zu verbinden sucht. Zu den Lernzielen gehörten die Auseinandersetzung mit fiktiven und nicht-fiktiven Geschichten von Vertriebenen und die Reflektion ihrer Lebensbedingungen im historischen Kontext. Im Rahmen der Lernaktivität wählten die Studierenden eine deutsche Stadt als Ausgangspunkt für eine historische Recherche zu digitalen Quellen über eine vertriebene Person oder Personengruppe nach dem Zweiten Weltkrieg. Auf der Basis der recherchierten und verifizierten Informationen entwickelten sie eine fiktive Figur und eine kurze Charakterbeschreibung mit Geburtsort, Kriegserfahrungen, Hindernisse der Rückkehr und daraus resultierende Schwierigkeiten. Die gesammelten Informationen wurden anschließend konsolidiert und mit Spielanweisungen verknüpft, um narrative Elemente für das Serious Game zu entwickeln. Angesichts der im Kapitel »AI NPCs & Immersive Learning« beschriebenen Herausforderungen hinsichtlich Authentizität und ethische Fragestellungen lag ein besonderer Fokus auf umfassendem Testen und Reflexion. Dabei testeten die Ler-

nenden ihre entworfenen KI-NPCs im Spiel, um deren Verhalten im Dialog zu erproben. Die Qualitätsbewertung der KI-NPCs erfolgte mithilfe des Godspeed Questionnaire (Bartneck 2023) und des Chatbot Usability Questionnaire (Holmes et al. 2023) und einer Abschlussreflexion. Durch die Einführung dieser standardisierten Bewertungsverfahren zur Beurteilung von Robotern und Chatbots erhielten die Lernenden wertvolle Anregungen, um KI-gesteuerte Spielfiguren in Bezug auf deren Wirkung und Verständlichkeit kritisch zu reflektieren.

Aus der Rolle fallen, um neue Perspektiven zu gewinnen

Das Ziel der Lernaktivität war in erster Linie die Auseinandersetzung mit dem Thema kriegsbedingte Migration und die Gefahren und Herausforderungen bei der Gestaltung von KI-NPCs.

Insgesamt führten die elf Lernenden während der Test- und Bewertungsaktivität 310 Dialogrunden mit ihren KI-NPCs durch, um Input zu identifizieren, das den KI-NPC dazu brachte, aus der Rolle zu fallen. Als Beispiele wurden Anachronismen wie Fragen zu modernen Persönlichkeiten der Gesellschaft, unsinnige Phrasen, Deduktionsfragen und sogar sehr einfache Eingaben wie Begrüßungen genannt.

Die Rückmeldungen der Studierenden auf die Reflexionsfragen spiegeln die intendierten Lernergebnisse wider und zeigen, dass sowohl KI-affine als auch KI-kritische Studierende die Lernaktivität interessant und motivierend fanden. Dabei waren sie in der Lage die Stärken und Schwächen der entwickelten KI-NPCs realistisch einzuschätzen und mit der Qualität der eigenen Prompts in Verbindung zu bringen, insbesondere in Bezug auf Informationen zum Kontext, wie z.B. relevante Hintergrundinformationen oder historische Quellen. Gleichzeitig regte die Aktivität zur Auseinandersetzung mit der Situation von Flucht und Vertreibung betroffener Menschen und deren Herausforderungen an.

Ethische Überlegungen betrafen die Zuverlässigkeit von Informationen, die Rolle der Opfer und deren Leiden, respektvolle Repräsentation und die Gefahr unangemessener Darstellung. Einige Studierende hinterfragten die Nutzung von KI in diesem sensiblen Kontext aus ethischen Gründen. Die Rückmeldungen der Studierenden decken sich weitgehend mit den Hauptanliegen von Kozlovski & Makhortykh (2025), die auf die Ambivalenz solcher digitalen Figuren hinweisen und Leitlinien für eine reflektierte ethische Bewertung bieten. In den Antworten wurden die folgenden ihrer Prinzipien aufgegriffen: authentische Präsenz (z.B. historische und psychologische Genauigkeit), das Einverständnis Betroffener (z.B. die Frage der Zustimmung verstorbener Personen), Wertschätzung (z.B. detaillierte, einfühlsame Darstellung) und Transparenz (z.B. um die Gefahr der Überschätzung der Genauigkeit durch Spielende zu minimieren). Kozlovski & Makhortykh's fünftes Anliegen, das Risiko, dass digitale Figuren die Erinnerungskultur beschädigen,

indem sie Erfahrungen von Zeitzeugen verzerren, trivialisieren oder gar kommerzialisieren, wurde in den Reflexionen der Studierenden nur am Rande gestreift, mit dem Verweis auf Stereotypen. In zukünftigen Lehrveranstaltungen könnte dieser Aspekt gezielt durch einen Denkanstoß über die Gefahren und Risiken, die mit der Gestaltung von KI-NPCs in historischen Lernkontexten verbunden sind, unterstützt werden. Angesichts unserer Pilotergebnisse glauben wir, dass solche Lernaktivitäten im Rahmen kollaborativer Spieleentwicklung fruchtbare Diskussionen anregen, die die Selbstwirksamkeit als Designer verbessern und die KI-Kompetenz steigern können. Die Graveler-Plattform hat sich dabei als zugängliche Plattform mit Potential zur Skalierfähigkeit für das Lernen durch Design- und Co-Creation-Aktivitäten für Lehrende und Studierende im Hochschulbereich gezeigt.

Zukünftig planen wir, diese praxisorientierte Reflexionsübung in jedem Seminar zu integrieren, in dem Studierende die KI-NPC-Funktion in Graveler nutzen und durch iterative Testzyklen zu erweitern. Wir sind überzeugt, dass dies zu einem besseren funktionalen und ethischen Verständnis der Potenziale und Risiken dieser Technologie beiträgt.

Literatur

Arnab, Sylvester/Clarke, Samantha/Morini, Luca (2019): »Co-creativity through play and game design thinking«, in: Electronic Journal of E-Learning, 17(3), S. 184–198.

Bartneck, Christoph (2023): »Godspeed questionnaire series: Translations and usage«. in: Chris Krägeloh/Oleg N. Medvedev/Mohnsen Alyami (Hg.), International handbook of behavioral health assessment, Springer International Publishing, S. 1–35.

Bykova, Ksenia/Frank, Ulrike (2016): »Spielend Lernen: Anwendbarkeit des« Dysphagia Game zur Vermittlung von Wissen über Symptome, Folgen und den Umgang mit Dysphagie«, in: Stadie, Nicole, Spektrum Patholinguistik 9: Schwerpunktthema: Lauter Laute: Phonologische Verarbeitung und Lautwahrnehmung in der Sprachtherapie, 9, S. 235–242.

Clark, Lynda (2025): Why Charles Dickens would have made Great Expectations a videogame if he were writing today. The Conversation vom 24. Februar 2025. https://theconversation.com/why-charles-dickens-would-have-made-great-expectations-a-videogame-if-he-were-writing-today-249199 vom 24.02.2025.

Da Silva, Guilherme A./de Souza Ribeiro, Marcos W. (2021): Development of Non-Player Character with Believable Behavior: a systematic literature revie, in: Simpósio Brasileiro de Jogos e Entretenimento Digital (SBGames), S. 319–323.

Deckert, Ronald/Heymann, Felix/Metz, Maren (2018): »Game-based Learning as Education Method in the Digital Age: Experiences at the Highest Military Education Institution in Germany with Online and Offline Game Formats Related to Developing Competencies«. in: A. Altmann/B. Ebersberger/C. Mössenlechner/D. Wieser (Hg.): The Disruptive Power of Online Education: Challenges, Opportunities, Responses. Emerald Publishing Limited, S. 185–204.

Dillig, Marcella/Böhmer, Anselm/Mason, Jon/Bolaji, Stephen/Isso, Illie/Stanescu, Miron/Sahin, Hilal (2024): »A Segue From Search to Dialogue: Leveraging GenAI for Pre-Service Teacher Training«. in: R. Sharma/A. Bozkurt (Hg.), Transforming Education With Generative AI: Prompt Engineering and Synthetic Content Creation. (IGI Global Scientific Publishing, S. 22–55.

Gonzales, Wilkinson D. W./Shen, Daniel J./Yan, Aihua/Xie, Nina/Francisco, Maria L./Wong, Paulina P. Y. (2024): »AI NPCs in an Educational Metaverse: Evaluating the Effectiveness of Prompt Templates for Contextual Interactions«, in: Eric C. K. Cheng (Hg.), Asia Education Technology Symposium, Singapore: Springer Nature, S. 53–74.

Holmes, Samuel/Bond, RR/Moorhead, Anne/Zheng, Jane/Coates, Vivien/McTear, Michael (2023): »Towards validating a chatbot usability scale«, in: Aaron Marcus/Elizabeth Rosenzweig/Marcelo S. Soares (eds), Design, User Experience, and Usability. HCII 2023. Lecture Notes in Computer Science, vol 14033. Switzerland: Springer Nature. S. 321–339.

James, Alison/Nerantzi, Chrissi (Hg.) (2019): The power of play in higher education: Creativity in tertiary learning, Cham: Springer.

Kafai, Yasmin B./Burke, Quinn (2015): »Constructionist gaming: Understanding the benefits of making games for learning«, in: Educational psychologist, 50(4). S. 313–334.

Körner, Kevin/Fink, Esther/Steiner, Ann-Kathrin/Chupakhin, Jennifer/Peres da Silva, Glaucia (2024): »Graveler – Global Awareness Education Through Virtual Voyages«, in: Avo Schönbohm/Francesco Bellotti/Antonio Bucchiarone/Francesca de Rosa/Manuel Ninaus/Alf Wang/Vanissa Wanick/Dondio Pierpaolo (Eds.), Games and Learning Alliance. 13th International Conference, GALA 2024, Berlin, Germany, November 20–22, Proceedings. Cham: Springer, S. 156–166.

Kozlovski, Atay/Makhortykh, Mykola (2025): »Digital Dybbuks and virtual Golems: the ethics of digital duplicates in Holocaust testimony«, in: Memory, Mind & Media, 4, e28, 1.

Lombardi, Marilyn M./Oblinger, Diana G. (2007): Authentic learning for the 21st century: An overview, in: Educause learning initiative, S. 1–12.

Marone, Vittorio (2016): Playful constructivism: Making sense of digital games for learning and creativity through play, design, and participation, in: Journal For Virtual Worlds Research, 9(3).

Rüth, Marco/Kaspar, Kai (2024): »Games im Unterricht. Einbettung, Effekte und zukünftiger Einsatz«. in: Johannes König/Charlotte Hanisch/Petra Hanke/Thomas Hennemann/Kai Kaspar/Matthias Martens/Sarah Strauß (Hg.), Auf die Lehrperson und ihren Unterricht kommt es an. Waxmann, S. 127–146.

Schwing, Yannik (2025): »Spielen mit KI. Emotionen im Umgang mit KI-gesteuerten NPCs«, in: Christoph Bareither (Hg.), KI:Kultur. Wie Künstliche Intelligenz in den Alltag findet, S. 135–160.

Shorey, Shefaly/Ang, Emily/Yap, John/Ng, Esperanza D./Lau, Siew T./Chui, Chee K. (2019): »A virtual counseling application using artificial intelligence for communication skills training in nursing education: development study«, in: Journal of medical Internet research, 21(10). e14658.

Söbke, Heinrich/Reichelt, Maria (2018): Sewer rats in teaching action: An explorative field study on students' perception of a game-based learning app in graduate engineering education. arXiv preprint arXiv:1811.09776.

Steinmaurer, Alexander/Dengel, Andreas/Comanici, Mario/Buchner, Josef/Memminger, Josef/Gütl, Christian (2024): »Immersive learning in history education: Exploring the capabilities of virtual avatars and Large Language Models«. in: Jule M. Krüger/Daniela Pedrosa/Dennis Beck/Marie-Luce Bourguet/Andreas Dengel/Rami Ghannam/Alan Miller/Anasol Peña-Rios/Jonathon Richter (Hg.), International conference on immersive learning, Switzerland: Springer Nature, S. 363–374

Whitton, Nicole (2012): »The place of game-based learning in an age of austerity«, in: Electronic Journal of e-Learning, 10(2), S. 249–256.

Whitton, Nicole (2014): »Digital games and learning: Research and theory«, in: Routledge.

Wouters, Pieter/Van Nimwegen, C./Van Oostendorp, H./Van Der Spek, Erik D. (2013): »A meta-analysis of the cognitive and motivational effects of serious games«, in: Journal of Educational Psychology, 105(2), S. 249–265.

Yang, Yu-Fen/Tseng, Chiefen C./Lai, Siao-Ching (2025): »Collaborative digital game design and play to facilitate college students' creativity development«, in: Interactive Learning Environments, 33(2), S. 1250–1269.

Yuan, Victor/Miller, Alan/Oliver, Iain (2025): »Immersive Education with Historical Characters: Conversational MetaHuman Based on Large Language Model, Speech Recognition and Generation«, in: Immersive Learning Research-Academic,1(1), S. 62–66.

Analyse qualitativer Daten mit KI-Unterstützung

Evolution statt Revolution qualitativer Evaluationsmethoden in der Qualitätssicherung der Philipps-Universität Marburg[1]

Florian Engel und Tristan Dohnt

Im Rahmen qualitativer Evaluationen hat sich das Einsatzspektrum von KI seit der Veröffentlichung von Large Language Models (LLMs) signifikant erweitert. Ein solches Einsatzszenario wird im Artikel vorgestellt: Die Einbindung der KI zur Unterstützung der inhaltsanalytischen Datenauswertung von Gruppendiskussionen in der Qualitätssicherungsarbeit von Studiengängen der Universität Marburg. Im Zentrum steht hierbei ein von uns entwickeltes, umfangreiches Codememo[2], das als Basis für die inhaltsanalytische Auswertung durch Künstliche Intelligenz dient. Eine knappe Reliabilitätsuntersuchung soll die Ergebnisgüte prüfen. Hierfür wurden zwei identische Datensätze sowohl von einem Menschen als auch von der KI inhaltsanalytisch kodiert/kategorisiert. Die Ergebnisse werden anschließend diskutiert. Es zeigt sich, dass es sich lohnt, bestehende Auswertungssysteme evolutionär weiterzuentwickeln, anstatt sie für den Einsatz durch KI von Grund auf neu zu konstruieren.

Analysis of qualitative data with AI support. Evolution instead of revolution of qualitative evaluation methods in quality assurance at Philipps University Marburg
The range of applications of AI in qualitative evaluations has expanded significantly since the publication of LargeLanguageModels. One such application scenario is presented in this article: the integration of AI to support the content-analytical data analysis of group discussions in the quality assurance work of degree programs at the University of Marburg. The focus here is a comprehensive code memo we developed, which serves as the basis for the content-

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 Codememos sind schriftlich festgehaltene analytische Notizen bzgl. eines Codes bzw. einer Kategorie (je nach Benennungskonvetion). Das Codememo dient dazu, die Bedeutung des/der übergeordneten Codes/Kategorie zwischen mehreren Analysierenden zu definieren und zu anderen Codes/Kategorien abzugrenzen. Sie dokumentieren damit die interpretativen Entscheidungen der Forschenden und unterstützen eine konsistente, transparente und nachvollziehbare Codierung über den gesamten Analyseverlauf hinweg – sind also eine gemeinsam entwickelte Konvention. (Saldana 2021 beschreibt coding memos, ohne sie selbst zu verwenden, Kukartz & Rädier 2024).

analytical analysis by artificial intelligence. A brief reliability study is intended to assess the quality of the results. For this purpose, two identical data sets were coded/categorized using content analysis by both a human and an AI. The results are then discussed. It shows that it is worthwhile to develop existing evaluation systems in an evolutionary manner rather than constructing them from scratch for use by AI.

Einleitung

Evaluationen gelten auch in Deutschland seit Jahrzehnten als zunehmend eigenes Forschungs- und Anwendungsfeld (Stockmann 2022). Gleichzeitig bezieht sich diese an sich bereits diverse Disziplin in Erkenntnistheorie und Methodologie weiterhin in großem Maße auf die ihr nahestehenden Geistes- und Sozialwissenschaften (Bryda et al. 2023). So gilt hier wie dort die Nutzung computergestützter qualitativer Datenanalyse (vermittels dezidierter QDA-Software) als Standard für Ausbildung und Alltagsgeschäft (Kuckartz/Rädiker 2024). Diese etablierten Vorgehensweisen sehen sich durch neue Entwicklungen im Bereich automatisierter Analyse ausgesetzt. An der Universität Marburg wollen wir diesen Wandel bewusst gestalten und unser methodisches Vorgehen sukzessive mit Hilfe von KI (weiter-)entwickeln. In diesem Artikel stellen wir einen Ansatz zur Integration von KI in die Analyse qualitativer Evaluationsdaten dar und beschreiben den damit einhergehenden Wandel als Evolution bestehender Methoden.

Eingangs erläutern wir unser generelles Vorgehen im Rahmen des Marburger Modells der Qualitätssicherung in Studium und Lehre. Anschließend gehen wir auf die qualitative Analyse von Evaluationsgesprächen ohne KI-Unterstützung ein und heben die Dokumentation in Form eines ›Kategorienschemas‹ hervor. Den Hauptteil beginnen wir mit einer knappen Erläuterung der beschafften technischen Lösung. Anschließend steht das sogenannte ›Codememo‹ als Herzstück unseres Ansatzes im Fokus. Hierbei handelt es sich um eine umfangreiche Sammlung an Kategoriendefinitionen und Beispielen, die unsere KI-unterstützte Analyse erst ermöglichen. Im abschließenden Abschnitt des Hauptteils weisen wir aus, wie wir die Ähnlichkeit der Kodierungen von Mensch und Maschine überprüft haben, was in das abschließende Fazit hinüberführt.

Qualitätssicherung und qualitative Inhaltsanalyse an der Universität Marburg

Als Teil der Zentralverwaltung ist es Aufgabe unseres Referats, Studiengänge über den gesamten student life cycle mit Evaluationsangeboten zu versorgen. Evaluation wird an der Universität Marburg in besonderem Maße fragestellungsbezogen be-

trieben (Stockmann 2002). Das heißt, es werden kaum regelmäßige universitätsweite Erhebungen durchgeführt. Stattdessen verfolgen wir ein Beratungsmodell, in dem jeder Qualitätssicherungskreislauf mit Sondierungsgesprächen beginnt (angelehnt an den PDCA-Zirkel (Deming 1986)). Hiervon ausgehend werden Daten in enger Abstimmung mit den relevanten Akteur*innen erhoben, ausgewertet und diskutiert, sowie darauf basierend Maßnahmen entwickelt, die in Zukunft überprüft werden sollen (Heinrich, 2020, Stockmann 2006).[3]

Nach der Klärung der interessierenden Themen und Fragen sowie der Entscheidung für einen qualitativen Workshop, setzen wir diesen in einer zwei- (QualiCheckFeedback) bis vierstündigen (QualiCheckStudientag) Session um, in der Audiodaten sowie Stichworte als Ergebnissicherung auf Postern als Dokumentation des Gesagten dienen. Diese werden im Sinne der qualitativen Inhaltsanalyse nach Kuckartz und Rädiker (2024) analysiert.

Qualitative Inhaltsanalyse bedeutet, Daten anhand deduktiv und/oder induktiv gebildeter Hauptkategorien sukzessive und immer feiner zu analysieren. Zentral für das Vorgehen ist, dass bereits absolvierte Analyseschritte erneut durchgeführt werden, wenn neue Daten vorliegen oder sich das Kategoriensystem derart weiterentwickelt hat, dass eine neuerliche Überprüfung von Kategorien durch erneutes Codieren sinnvoll erscheint.

Im Gegensatz zur Methodenumsetzung mit wissenschaftlichem Anspruch, vor allem eines iterativen Auswertungsprozesses, besteht die Datenerhebung und -auswertung an der Universität Marburg jeweils aus einem zusammenhängenden Schritt und dauert in aller Regel höchstens sechs Monate von der Instrumentenabsprache (über die Bewerbung und Durchführung des Workshops) bis zum fertigen Bericht. Gerade der Aspekt der Überprüfung von Daten und des Zurücktretens im Prozess – umhin also die langfristige Weiterentwicklung und Verbesserung der Kategorien, ist in unserer alltäglichen Arbeit nur eingeschränkt möglich.

Als Antwort darauf haben wir vor einer Weile begonnen, die reichlich vorhandenen Berichte vergangener qualitativer Evaluationsmaßnahmen zu sammeln, zu sortieren und zu kategorisieren. Das Ergebnis ist eine sehr umfangreiche Excel-Tabelle, die uns als Dokumentation und Erinnerungsanker dafür dient, welche Haupt- und Subkategorien aus welchen Daten über die Jahre entwickelt wurden. Mit dieser als ›Kategorienschema‹ benannten Datei lassen sich neu erhobene Datensätze in Bezug zu vergangenen Evaluationsdurchläufen setzen und auswerten. Durch die fortlaufende Implementation neuer Auswertungsberichte aktualisieren wir zugleich diese Datenbasis des ›Kategorienschemas‹ und erfüllen damit ›in the long run‹ den Anspruch des kuckartz'schen Vorgehens, in der Analyse an bereits erfolgte

3 Für eine ausführliche, aktuelle Beschreibung des Marburger Ansatzes siehe Dohnt/Engel 2025.

Arbeitsschritte zurückzukehren (Kuckartz/Rädiker 2024: 142). Tatsächlich überprüfen wir damit nicht nur neue Daten, sondern modifizieren und verbessern unser Kategoriensystem fortwährend.

Qualitative Analyse von Evaluationsdaten mit generativer KI

Technische Basis

Für die Umsetzung unserer KI-Analysen haben wir das Modul AI Assist für das Programm MaxQDA der Firma Verbi beschafft. Dieses nutzt zum Zeitpunkt der Studie im Juli 2025 das LLM ChatGPT von OpenAI in der Version 3.5. AI Assist bietet verschiedene Funktionen zur Nutzung generativer KI im Arbeitsprozess. Im Rahmen erster Tests erwies sich die Funktion mit multiplen Dokumenten zu chatten als sinnvollste Variante. Während andere Funktionen einzelne Kategorien automatisch zusammenfassen oder analysieren, wird im Dokumenten-Chat ein separates Chatfenster im User*inneninterface geöffnet, über das mit allen ausgewählten Dateien (auch dem Codememo) zugleich gearbeitet werden kann, was sehr ähnlich zur Bedienung in Schnittstellen via Browser funktioniert. Eine hier angewiesene, automatische Analyse wird so verarbeitet, als würde es sich um ein einzelnes Dokument handeln. Zugleich besteht keine Möglichkeit, weitere Einstellungen bzgl. Kreativität, Logik, Fokus oder Diversität der Antworten der eingesetzten KI vorzunehmen, wie in gängigen Browserschnittstellen – was angesichts der Nachvollzieh- und Reproduzierbarkeit durchaus sinnvoll erscheint.

Codememo

Von zentraler Bedeutung unseres Ansatzes ist – wie bereits angedeutet – das sogenannte ›Codememo‹. Ein Codememo stellt in der Anwendung von QDA-Software durch einen Menschen eine kurze Zusammenfassung dessen dar, was eine Kategorie bedeuten und umfassen soll und ist im User*inneninterface auch mit diesem verbunden (Kuckartz/Rädiker 2024: 122–123, 208–214). Dies ermöglicht es menschlichen Anwender*innen, ein gemeinsames Verständnis einer (Sub-)Kategorie zu entwickeln, was die Analyse und ihre Ergebnisse besser und nachvollziehbarer gestaltet. Auch ›unser‹ Codememo versammelt Definitionen und Beispiele für existierende Haupt- und Subkategorien, die wir aus der deutlich umfangreicheren Kategorienschema-Datei entwickelt haben. Diese Excel-Datei wurde hierzu sowohl händisch wie auch durch KI zusammengefasst, strukturiert und neu formatiert. Es umfasst in seiner aktuellen Version 12 Oberkategorien, 92 Unterkategorien sowie knappe Analysehinweise genereller Art mit insgesamt knapp 65.000 Zeichen in ei-

ner einzelnen Word-Datei. Diese Datei durfte aufgrund der Zeichenbegrenzungen von AI Assist im Juli 2025 100.000 Zeichen nicht überschreiten.

Die Definitionen jeder Hauptkategorie werden im Codememo nach demselben Muster festgehalten:

1. Definitionen Oberkategorie (erläuternder Fließtext plus Beispiele)
2. Definitionen Subkategorien (erläuternder Text plus Beispiele in bullet point-Struktur)
3. Analysehinweise genereller Art (erläuternder Text in bullet point-Struktur)

Dieses Codememo erfüllt also einen vergleichbaren Zweck wie das gleichnamige Element im gängigen analytischen Vorgehen menschlicher Evaluator*innen und Sozialforscher*innen (Kuckartz/Rädiker 2024: 208–214). Statt eine Konvention unter menschlichen Analyst*innen zu sein, liefert es einer generativen KI die grundlegenden Informationen dazu, anhand welcher Kategorien und Definitionen »sie« vorgegebene Daten zu analysieren hat. Der genutzte Prompt enthält neben basalen Verarbeitungsanweisungen auch Kontextinformationen zu den Daten, den zentralen Verarbeitungsauftrag sowie Ausgabeanweisungen. Letztere dienen dazu, eine stabile, in unsere Berichtsform übertragbare Ergebnisstruktur zu erhalten. In Zukunft sollen auch diese Ergebnisse zur Überarbeitung und Weiterentwicklung des Kategorienschemas und damit des Codememos herangezogen werden.

Vergleich der Kodierungen von Mensch und Maschine

Im Rahmen des Entwicklungsprozesses haben wir Daten aus zwei verschiedenen Evaluationsmaßnahmen (einer Modul- und einer Studiengangsevaluation) sowohl von einem Menschen als auch mit unserem KI-Verfahren inklusive Codememo auswerten lassen. Die Analyse umfasste insgesamt 75 Zitatstellen. Mensch und KI ordneten in 74,14 % der Fälle die Zitate derselben Oberkategorie und in 87,88 % der Fälle derselben Unterkategorie zu. Somit war die Übereinstimmung zwischen Mensch und KI signifikant höher als der Zufall (Oberkategorie: $t(57)=4{,}16$, $p<.001$; Unterkategorie: $t(65)=9{,}36$, $p<.001$)[4]. Die Kodierungen wurden von zwei unabhängigen Raterinnen durchgeführt. Die Interrater-Reliabilität lag bei der Oberkategorie bei Cohen's Kappa $\kappa = 0{,}486$ und bei der Unterkategorie bei Cohen's Kappa $\kappa = 0{,}575$, was eine mäßige, aber signifikante Übereinstimmung zwischen den Raterinnen darstellt und somit als verlässlich angesehen werden kann.

Für den ersten Test sind wir mit diesen Ergebnissen zufrieden. Sie legen nahe, dass die entwickelte automatische Auswertungsmethode Ergebnisse liefert, die mit

4 Für die Vergleiche und Signifikanztests wurden nur die Fälle berücksichtigt, in denen die beiden Raterinnen in ihren Urteilen übereingestimmt haben.

Ergebnissen unserer eigenen Analysen vergleichbar genug sind, um als Basis für die künftige Weiterentwicklung zu dienen, ähnlich den Erkenntnissen von Gao et al. (2023).

Fazit

Wie im letzten Abschnitt zum Reliabilitätsmaß bereits angedeutet wurde, haben sich das Codememo und die Vorgehensweise der Auswertung bis zu diesem Zeitpunkt bereits bewährt. Der unausweichlichen Voreingenommenheit menschlicher Rater*innen wurde zwar durch Beispiele für das Rating seitens der Autor*innen begegnet – menschliche Bewertung bleibt letzten Endes aber ein subjektiver Prozess. »Opak« (Schmitt/Heckwolf 2024: 53) bleibt aber auch, was im Prozess der »tatsächliche[n] Verarbeitung der Daten« einer mit Deep Learning trainierten KI geschieht. Was jedoch bereits an der Oberfläche von KI produzierten Textes auffällt und sich auch in der Literatur findet, ist, dass KI-Analyse (weiterhin) Kontextinformationen fehlen bzw. sie diese nicht (sicher) erfasst, was Ergebnisse oberflächlicher Zuordnungen weniger nachvollziehbar und Schlussfolgerungen teils widersprüchlich wirken lässt (Roberts et al. 2023; Mostafapour et al. 2024; Thelwall 2025).

Verbesserungen erscheinen dennoch möglich. Diese konzentrieren sich derzeit vornehmlich auf das Codememo als zentrales analytisches Raster. Dieses sollte vor der nächsten Anwendung um die aktuellen Ergebnisse ergänzt werden. Gerade hier fehlt uns aber noch ein erprobtes Vorgehen, das nicht den gesamten Entwicklungsprozess neu aufrollt. Schon aus dieser Sicht handelt es sich bei unserem Ansatz um einen evolutionären, insofern wir neue Prozesse entlang des Weges entwickeln und bei Erfolg implementieren. Verbesserung ist aber auch im Prozessablauf nötig und möglich. Andere Untersuchungen verweisen zusätzlich darauf, dass Prompts ein beständiges Feld der Unsicherheit und Verbesserung darstellen (Fuller et al. 2024; Mostafapour et al. 2024) – was wir einerseits bestätigen können, andererseits nicht die zentrale Herausforderung darstellt.

Das analytische Vorgehen – mit dem existierenden Codeschema – innerhalb des QDA-Programms ist dagegen derart routinisiert, dass es sich bei kommenden Evaluationsmaßnahmen bereits anbietet, menschliche und automatisch erstellte Auswertung quasi parallel durchzuführen, inhaltlich zu vergleichen und die menschliche Version durch die KI-erstellte zu ergänzen. Ideal wäre es allerdings zugleich, erneute Reliabilitätsüberprüfungen in den Ablauf zu integrieren. Hieran erkennt man wiederum die Limitationen einer Entwicklung entlang des Weges, da sich Überprüfungen und Tests nicht immer in das Alltagsgeschäft einbinden lassen.

Literatur

Bryda, G./Costa, A. P. (2023): Qualitative Research in Digital Era: Innovations, Methodologies and Collaborations, in: Social Sciences, 12 (10), 570.

Deming, W. E. (1986): Out of the Crisis, Cambridge: MIT Press.

Dohnt, Tristan/Engel, Florian (2025): »Qualitätssicherung als per se ko-kreativer Prozess. Die Marburger Qualitätssicherung für Studium und Lehre«. in: Jörg Jörissen/Daniela Möller/Martina Grein/Sacha Kopczynski/David Peters (Hg.), Evaluation von Studium und Lehre ko-kreativ gestalten, Münster/New York: Waxmann, S. 101–108.

Heinrich, Mathis (2020): »Nachhaltigkeit durch Flexibilität, Reichweite und Diffusion: Das Marburger Modell einer Qualitätssicherung »on demand«. in: QiW – Qualität in der Wissenschaft, 14(1), 17–26.

Kuckartz, Udo/Rädiker, Stefan (2024): Qualitative Inhaltsanalyse. Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz, Weinheim: Beltz Juventa.

Roberts, Richard Hr./Ali, Stephen R./Hutchings, Hayley A./Dobbs, Thomas D./Whitaker, Ian S. (2023): »Comparative study of ChatGPT and human evaluators on the assessment of medical literature according to recognised reporting standards«, in: BMJ Health and Care Informatics, 30 (1).

Saldana, Johnny (2021): The Coding Manual for Qualitative Researchers, Los Angeles/London/New Dehli/Singapore/Washington DC: SAGE

Schmitt, Marco/Heckwolf, Christoph (2024): »KI zwischen Blackbox und Transparenz:«, in: Roger Häußling/Claudius Härpfer/Marco Schmitt (Hg), Soziologie der Künstlichen Intelligenz: Perspektiven der Relationalen Soziologie und Netzwerkforschung, Bielefeld: transcript Verlag, S. 51–84.

Stockmann, Reinhard (2002): Was ist eine gute Evaluation. CEval-Arbeitspapiere, Nr. 9, Centrum für Evaluation. PID: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-118018

Stockmann, Reinhard (2006): Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement. (Bd. 5), Münster/New York: Waxmann.

Stockmann, Reinhard (Hg.) (2022): Handbuch zur Evaluation. Eine praktische Handlungsanleitung. (Band 16), Münster/New York: Waxmann.

Thelwall, Mike (2025): »ChatGPT for complex text evaluation tasks«, in: Journal of the Association for Information Science and Technology, 76 (4), S. 645-648.

KI in der Lehre – Eine aktuelle Übersicht über die Nutzung von KI im Studienkontext[1]

Rischert Enes[2], Arnim Spengler[3], Felix Meckmann[4], Thanh Viven Le-Vu[5] und Khaoula Benmaarouf[6]

Das Projekt KI4Edu untersucht den Einsatz von KI-Technologien in der Lehre des Bauwesens, der Immobilienwirtschaft und der Energiewirtschaft. Eine deutschlandweite Studie wurde durchgeführt, um die aktuelle Nutzung von KI im Studienkontext zu erfassen und innovative Lehrmethoden zu entwickeln. Die Umfrage analysiert die Nutzungshäufigkeit von KI-Tools, deren Einfluss auf die akademische Leistung sowie Erwartungen, Bedenken und Vorteile. Erste Ergebnisse zeigen, dass über 90 % der Befragten KI-Anwendungen im Studium einsetzen, was zu positiven Effekten in Effizienz und Qualität der Lehre führt. Darüber hinaus befasst sich das Projekt mit dem praktischen Einsatz von Large Language Models (LLMs) in der Hochschullehre. Im Fokus stehen deren Potenziale zur Unterstützung bei Recherche, Textproduktion und Strukturierung von Lerninhalten sowie die Grenzen in Bezug auf fachliche Tiefe, didaktische Eignung und Datenschutz. Um diese Herausforderungen zu adressieren, wurde ein prototypischer Ansatz entwickelt, der LLMs über eine Retrieval-Augmented-Generation-(RAG)-Architektur mit kuratierten Wissensbasen verknüpft. Dadurch sollen Genauigkeit, Nachvollziehbarkeit und fachliche Relevanz der generierten Antworten verbessert werden. Die Ergebnisse bilden eine Grundlage für die Entwicklung praxisorientierter Konzepte zur Integration von KI in die Hochschullehre und leisten einen Beitrag zur digitalen Transformation des Bildungsbereichs.

AI in Teaching – A Current Overview of the Use of AI in the Study Context

The KI4Edu project examines the application of AI technologies in civil engineering, real estate, and energy economics education. A nationwide study was conducted to map the current use of AI in academic settings and foster innovative teaching methods. The survey assesses AI tool

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0009–0003-3851-1287
3 ORCID: 0000–0003-0920-5523
4 ORCID: 0000–0003-0990-4789
5 ORCID: 0009–0008-9423-1526
6 ORCID: 0009–0006-5234-5766

usage, its impact on academic performance, and student expectations, concerns, and benefits. Initial findings indicate that over 90 % of respondents integrate AI applications into their studies, yielding positive effects on teaching efficiency and quality. In addition, the project addresses the practical use of Large Language Models (LLMs) in higher education. The focus lies on their potential to support research, text generation, and the structuring of learning materials, while also highlighting limitations regarding subject-specific depth, didactic suitability, and data protection. To address these challenges, a prototype has been developed that links LLMs with curated knowledge bases using a Retrieval-Augmented Generation (RAG) architecture. This approach aims to enhance accuracy, traceability, and subject relevance of generated responses. The findings provide a foundation for developing practical concepts to integrate AI into higher education and contribute to the digital transformation of the education sector.

Einleitung und Problemstellung

Generative KI-Systeme und Large Language Models (LLMs) werden von Studierenden bereits breit zur Lernunterstützung eingesetzt, insbesondere zur Erklärung von Fachinhalten und zur Recherche. Gleichzeitig bestehen substanzielle Vorbehalte hinsichtlich der inhaltlichen Verlässlichkeit, möglicher Halluzinationen sowie didaktischer und datenschutzrechtlicher Aspekte. Der Mehrwert von LLMs in der Hochschullehre hängt daher maßgeblich von ihrer kontrollierten, transparenten und didaktisch fundierten Integration ab.

Das vom Bundesministerium für Bildung und Forschung geförderte Projekt KI4Edu untersucht die Nutzung und Bewertung von KI-Technologien im Studium auf Basis einer deutschlandweiten Studierendenbefragung. Die Ergebnisse liefern empirische Hinweise auf zentrale Einsatzbereiche, Erwartungen und Herausforderungen und bilden damit eine Grundlage für die Ableitung konkreter Anforderungen an KI-gestützte Lernsysteme.

Der vorliegende Beitrag greift diese Befunde auf und leitet daraus Anforderungen an den Einsatz von LLMs in der Hochschullehre ab. Vorgestellt wird ein prototypischer Ansatz, der LLMs mittels Retrieval-Augmented Generation (RAG) mit einer kuratierten Wissensbasis kombiniert, um Transparenz, Nachvollziehbarkeit und inhaltliche Zuverlässigkeit systematisch zu verbessern. Ziel ist die konzeptionelle und technische Ausgestaltung von LLMs als lernunterstützende Systeme im akademischen Kontext.

Abgrenzung und Begriffsdefinition

Im vorliegenden Beitrag wird der Begriff der Künstlichen Intelligenz (KI) auf generative, dialogbasierte Sprachmodelle beschränkt. KI bezeichnet allgemein rechnergestützte Systeme zur Ausführung kognitiver Aufgaben; im Zentrum der Betrachtung stehen hier jedoch Large Language Models (LLMs) als Teilmenge der generativen KI, die auf großen Textkorpora trainiert werden und kontextbezogene Texte generieren können. Der Fokus liegt ausschließlich auf dem Einsatz solcher LLM-basierter Systeme in der Hochschullehre; andere KI-Anwendungen, etwa in Planung, Optimierung oder Administration, werden nicht betrachtet, da sich deren Nutzungsszenarien und didaktische Implikationen grundlegend unterscheiden. Zudem wird zwischen der informellen, individuellen Nutzung von LLMs durch Studierende und ihrem institutionell verantworteten Einsatz in Lehrveranstaltungen unterschieden, der zusätzliche Anforderungen an Transparenz, Nachvollziehbarkeit, fachliche Korrektheit und Datenschutz stellt. Diese begriffliche und konzeptionelle Eingrenzung bildet die Grundlage für die folgende Analyse der empirischen Ergebnisse und die Ableitung didaktischer sowie technischer Anforderungen.

Ergebnisse der KI4Edu Studie

Die empirischen Befunde des Beitrags basieren auf einer deutschlandweiten Online-Befragung von Studierenden im Rahmen des Projekts KI4Edu, die zwischen Juli und Oktober 2024 durchgeführt wurde. Nach Datenbereinigung gingen 1.020 gültige Fragebögen von Studierenden aus 85 Hochschulen in die Auswertung ein. Die Ergebnisse zeigen eine sehr hohe Verbreitung von KI-Technologien im Studienalltag: Über 90 % der Befragten nutzen KI regelmäßig, wobei Chatbots auf Basis von Large Language Models die mit Abstand dominierende Anwendung darstellen. LLMs werden vor allem zur Erklärung von Fachinhalten, zur Recherche und zur Unterstützung beim Verständnis komplexer Themen eingesetzt. Gleichzeitig differenzieren Studierende klar zwischen unterstützenden und bewertenden Einsatzszenarien. Während KI-gestützte Lernunterstützung überwiegend positiv bewertet wird, besteht gegenüber automatisierten Prüfungs- und Bewertungsformaten eine ausgeprägte Skepsis. Als zentrale Bedenken werden insbesondere die Verlässlichkeit der generierten Inhalte, mögliche Halluzinationen, datenschutzrechtliche Unsicherheiten sowie eine potenzielle Abhängigkeit von KI-Systemen genannt. Auffällig ist zudem, dass viele Studierende den Einsatz von KI bislang kaum im Rahmen formaler Lehrveranstaltungen erlebt haben, was auf eine Diskrepanz zwischen individueller Nutzung und institutioneller Integration hinweist. Insgesamt verdeutlichen die Ergebnisse eine ambivalente Ausgangslage: LLMs

sind im Studienalltag etabliert und werden als lernförderlich wahrgenommen, zugleich bestehen klare Anforderungen an Transparenz, Nachvollziehbarkeit und eine didaktisch verantwortete Einbettung. Diese Befunde bilden die empirische Grundlage für die im weiteren Verlauf abgeleiteten didaktischen und technischen Anforderungen sowie für die Konzeption eines prototypischen LLM-gestützten Lernsystems. (Enes, Meckmann, & Le-Vu, 2025)

Didaktische Implikationen aus den Studienergebnissen

Die Ergebnisse der KI4Edu-Studie lassen sich in zentrale didaktische Anforderungen an den Einsatz von Large Language Models in der Hochschullehre überführen. Studierende akzeptieren LLMs primär als lernunterstützende Werkzeuge zur Erklärung, Strukturierung und Vertiefung von Inhalten, während ihr Einsatz in bewertenden Kontexten auf deutliche Skepsis stößt. Daraus folgt die Notwendigkeit einer klaren Trennung zwischen lernunterstützenden Funktionen und prüfungsrelevanten Formaten. Zugleich zeigen die Befunde, dass Vertrauen in LLMs maßgeblich von Transparenz und Nachvollziehbarkeit abhängt, insbesondere im Umgang mit potenziell fehlerhaften oder halluzinierten Antworten. Didaktisch erfordert dies Systeme, die Unsicherheiten sichtbar machen, Quellen offenlegen und Studierende zur kritischen Reflexion anregen. Darüber hinaus wird deutlich, dass der Einsatz von LLMs nicht auf reine Tool-Nutzung reduziert werden darf, sondern gezielt den Aufbau von KI-Kompetenzen unterstützen sollte, etwa durch die kritische Bewertung und Weiterentwicklung KI-generierter Inhalte sowie die Auseinandersetzung mit Datenschutz- und Verantwortungsfragen. Schließlich verweist die geringe Sichtbarkeit von KI in formalen Lehrveranstaltungen auf die Notwendigkeit einer institutionellen und curricularen Einbettung. Lehrende übernehmen hierbei eine zentrale Rolle, indem sie Einsatzszenarien definieren und den Einsatz von LLMs didaktisch rahmen. Insgesamt ergeben sich Anforderungen an LLM-gestützte Lernsysteme, die Lernunterstützung priorisieren, Transparenz gewährleisten, Kompetenzaufbau fördern und systematisch in bestehende Lehrkonzepte integriert sind. Diese Anforderungen bilden die Grundlage für die nachfolgende Konzeption eines prototypischen LLM-gestützten Lernsystems

Konzeption eines LLM-gestützten Lernsystems (Prototyp)

Auf Grundlage der empirischen Ergebnisse der KI4Edu-Studie wurde ein prototypisches, LLM-gestütztes Lernsystem konzipiert, das generative Sprachmodelle gezielt zur Unterstützung individueller Lernprozesse in der Hochschullehre einsetzt. Der Ansatz folgt klaren didaktischen Leitprinzipien: LLMs werden als lernunter-

stützende Assistenzsysteme verstanden, nicht als Instrumente zur Leistungsbewertung; Transparenz und Nachvollziehbarkeit der generierten Inhalte haben Vorrang; zugleich soll der kritische und kompetente Umgang mit KI aktiv gefördert werden. Zur Umsetzung dieser Anforderungen wird ein Retrieval-Augmented-Generation-Ansatz verwendet, bei dem ein LLM mit einer kuratierten, fachlich geprüften Wissensbasis kombiniert wird. Relevante Lehr- und Lernmaterialien werden strukturiert aufbereitet, semantisch erschlossen und kontextabhängig in den Antwortgenerierungsprozess eingebunden. Dadurch wird die Abhängigkeit vom vortrainierten Weltwissen reduziert, die Nachvollziehbarkeit erhöht und das Risiko inhaltlicher Fehlannahmen verringert. Ergänzend adressiert die Systemkonzeption Fragen von Transparenz und Datenschutz durch flexible Einsatzszenarien sowie durch die explizite Kommunikation von Systemgrenzen und Abhängigkeiten von der Qualität der Wissensbasis. Das Lernsystem ist modular aufgebaut und erlaubt eine curriculare Anpassung an unterschiedliche Lehrveranstaltungen. Es versteht sich als Ergänzung bestehender Lehrformate, insbesondere für selbstgesteuerte Lernphasen, Wiederholung und Vertiefung, während didaktische Steuerung und Bewertung weiterhin bei den Lehrenden verbleiben. Insgesamt operationalisiert der Prototyp die aus der Studie abgeleiteten Anforderungen an lernunterstützende, transparente und verantwortungsvoll eingebettete LLM-Systeme in der Hochschullehre.

Diskussion

Die Ergebnisse der KI4Edu-Studie und der darauf aufbauende prototypische Ansatz verdeutlichen, dass der Einsatz von Large Language Models in der Hochschullehre nur dann einen nachhaltigen didaktischen Mehrwert entfalten kann, wenn er konzeptionell und institutionell reflektiert erfolgt. Die empirischen Befunde zeigen eine hohe Akzeptanz generativer KI im individuellen Studienalltag, zugleich jedoch eine deutliche Zurückhaltung gegenüber ihrem formalen Einsatz in Lehrveranstaltungen und insbesondere in prüfungsrelevanten Kontexten. Der vorgestellte Ansatz greift diese Diskrepanz auf, indem er LLMs konsequent als lernunterstützende Systeme konzipiert und ihre Rolle klar von Bewertungs- und Entscheidungsfunktionen abgrenzt. Damit fügt er sich in hochschuldidaktische Konzepte ein, die selbstgesteuertes Lernen, formative Unterstützung und die Förderung metakognitiver Kompetenzen betonen. Der Einsatz von Retrieval-Augmented Generation stellt dabei einen wichtigen Beitrag zur Erhöhung von Transparenz und Nachvollziehbarkeit dar, indem generative Antworten an kuratierte, fachlich geprüfte Inhalte gebunden werden. Zugleich bleiben systemimmanente Unsicherheiten bestehen, die sowohl von der Qualität der Wissensbasis als auch von interpretativen Spielräumen des Modells abhängen. Diese Grenzen machen deutlich, dass RAG kein technisches Allheilmittel darstellt, sondern didak-

tisch begleitet und reflektiert eingesetzt werden muss. Eine zentrale Rolle kommt dabei den Lehrenden zu, die sowohl die inhaltliche Ausgestaltung der Wissensbasis als auch die didaktische Rahmung verantworten. Der erfolgreiche Einsatz LLM-gestützter Lernsysteme erfordert daher organisatorische, technische und institutionelle Rahmenbedingungen, etwa zeitliche Ressourcen, Supportstrukturen und klare hochschulische Leitlinien. Der prototypische Ansatz weist aufgrund seiner modularen Architektur eine grundsätzliche Übertragbarkeit auf unterschiedliche Fachdisziplinen auf, bleibt jedoch in seinen Möglichkeiten begrenzt und kann insbesondere in forschungs- oder diskursorientierten Lehrformaten nur unterstützend wirken. Insgesamt unterstreicht die Diskussion, dass der didaktische Mehrwert von LLMs weniger in ihrer technischen Leistungsfähigkeit als in ihrer reflektierten Einbettung in Lehr- und Lernkontexte liegt und eine kontinuierliche Auseinandersetzung mit didaktischen, organisatorischen und ethischen Fragen erfordert.

Zusammenfassung und Ausblick

Die Ergebnisse der KI4Edu-Studie belegen eine hohe Verbreitung und Akzeptanz von KI-Tools im Studienalltag, machen jedoch zugleich deutliche Vorbehalte gegenüber ihrem Einsatz in bewertenden Kontexten sowie Unsicherheiten hinsichtlich Transparenz, Verlässlichkeit und Datenschutz sichtbar.

Ausgehend von diesen Befunden wurden zentrale didaktische Anforderungen abgeleitet und in die Konzeption eines prototypischen LLM-gestützten Lernsystems überführt. Der vorgestellte Ansatz versteht LLMs konsequent als lernunterstützende Assistenzsysteme und grenzt sie klar von prüfungsrelevanten Entscheidungsfunktionen ab. Durch die Kombination generativer Sprachmodelle mit einer kuratierten Wissensbasis mittels Retrieval-Augmented Generation wird versucht, insbesondere dem Halluzinationsproblem und der mangelnden Nachvollziehbarkeit systematisch zu begegnen.

Der Beitrag leistet damit einen doppelten Beitrag: Einerseits wird aufgezeigt, wie empirische Erkenntnisse zur KI-Nutzung von Studierenden in konkrete didaktische Anforderungen übersetzt werden können. Andererseits wird ein konzeptioneller Ansatz vorgestellt, der technische Architekturentscheidungen explizit aus didaktischen Zielsetzungen ableitet und so eine Verbindung zwischen empirischer Bildungsforschung und technischer Systementwicklung herstellt. Zugleich wird deutlich, dass LLM-gestützte Lernsysteme weder Lehrende ersetzen noch diskursive Lehrformate substituieren können, sondern ihre Stärken insbesondere in der Unterstützung selbstgesteuerter Lernprozesse sowie in der Wiederholung und Vertiefung von Inhalten liegen.

Für die weitere Forschung besteht insbesondere Bedarf an empirischen Evaluationen des vorgestellten Prototyps in realen Lehr-Lern-Szenarien, um dessen didaktischen Nutzen, Akzeptanz und Grenzen systematisch zu untersuchen. Darüber hinaus ist die Übertragbarkeit auf andere Fachdisziplinen sowie das Potenzial von LLMs zur gezielten Förderung kritischen Denkens weiter zu analysieren. Insgesamt verdeutlicht der Beitrag, dass der didaktisch sinnvolle Einsatz von LLMs weniger von technologischer Leistungsfähigkeit als von reflektierter Konzeption, empirischer Fundierung und institutioneller Unterstützung abhängt.

Literatur

Enes, Rischert/Meckmann, Felix/Le-Vu, Thanh V. (2025): KI in der Lehre – Eine aktuelle Übersicht über die Nutzung von KI im Studienkontext. https://doi.org/10.71764/opus-1576

Systematische Evaluation prompt-basierter, durch KI generierte Multiple-Choice-Fragen in der Medizinischen Biochemie[1]

Masoud Darabi[2] und Shabnam Fayezi[3]

Künstliche Intelligenz (KI)-Sprachmodelle eröffnen neue Möglichkeiten in der medizinischen Ausbildung und können u. a. zur schnellen Erstellung von Multiple-Choice-Fragen (MCQs) dienen. Diese Studie bewertete prompt-basierte, KI-generierte MCQs im Vergleich zu Dozierenden-generierten Fragen in der medizinischen Biochemie hinsichtlich Qualität, kognitivem Niveau und klinischer Relevanz. Hierzu erzeugten drei KI-Plattformen jeweils 20 MCQs zu Stoffwechsel und Krankheitsprozessen. Ergänzt um 20 von Dozierenden erstellte Fragen wurden alle Fragen von fünf Expert:innen anhand eines strukturierten Rasters bewertet. ChatGPT-4 schnitt insgesamt am besten ab; Gemini 2.5 Pro lieferte klinisch relevantere Fragen; Copilot bot die klarsten Vorlagen. Häufige Schwächen waren mehrdeutige Fragestellungen und redundante Formulierungen. Dieser Ansatz ermöglicht effiziente Fragengenerierung, wobei Stärken und Grenzen der jeweiligen Plattform die Auswahl je nach Bildungsziel beeinflussen.

Systematic Evaluation of prompt-based AI-generated multiple-choice questions in medical biochemistry

Artificial intelligence (AI) language models offer new opportunities in medical education, including rapid generation of multiple-choice questions (MCQs). This study evaluated prompt-based AI-generated MCQs in medical biochemistry against instructor-designed items for quality, cognitive level, and clinical relevance. Three AI platforms each generated 20 questions on metabolism and disease processes. Five experts assessed MCQs using a structured rubric. ChatGPT-4 performed best overall; Gemini 2.5 Pro produced more clinically relevant MCQs; Copilot provided the clearest templates. Common weaknesses included ambiguous questions and redundant wording. This approach enables efficient MCQ generation, and platform-specific strengths and limitations influence selection depending on the educational objective.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000–0001-6380-272X

3 ORCID: 0000–0001-6954-5138

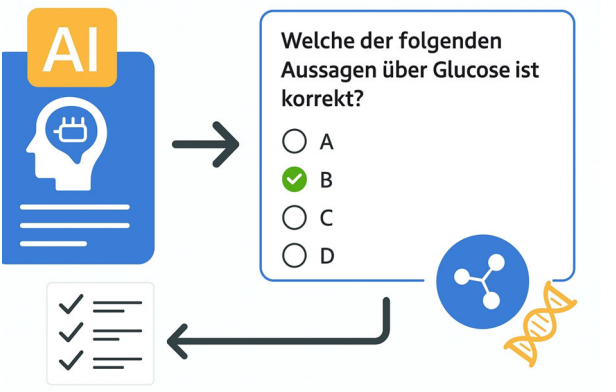
Einleitung

Künstliche Intelligenz (KI), insbesondere große Sprachmodelle (Large Language Models, LLMs), gewinnt in der medizinischen Ausbildung rasch an Bedeutung und verändert vor allem Lehr- und Prüfungsprozesse sowie die Erstellung von Lernmaterialien. Zu den meistgenannten Potenzialen zählt die Möglichkeit, innerhalb kürzester Zeit eine große Anzahl vielfältiger Prüfungsfragen zu generieren und damit die Arbeitsbelastung von Lehrenden zu reduzieren (Abd-alrazaq et al., 2023).

Aktuelle Studien untersuchen den Einsatz von LLMs zur automatisierten Erstellung von Multiple-Choice-Fragen (MCQs), häufig unter Verwendung von Retrieval Augmented Generation (Grévisse et al., 2024; Hang et al., 2024; Kunuku & Dehbozorgi, 2025). Dabei werden, um Qualität und Verständlichkeit zu bewerten, Fragen aus Lehrtexten und Kursmaterialien abgeleitet und anschließend mit manuell erstellten Fragen verglichen. Im Gegensatz dazu basiert die Prompt-basierte, auch als »Closed-Book« bezeichnete, Generierung ausschließlich auf den in den Modellparametern aus den Trainingsdaten abgeleiteten Informationen, ohne externe Quellen abzurufen. Beide Ansätze zeigen Potenzial, unterscheiden sich jedoch hinsichtlich Nachvollziehbarkeit und curricularer Anwendbarkeit. Moderne LLMs sind auf umfangreichen Datensätzen trainiert, was wiederum eine hohe Leistungsfähigkeit in der Sprachverarbeitung sowie die Bearbeitung fachspezifischer Aufgaben, etwa im medizinischen Bereich, ermöglicht. Daher gewinnen prompt-basierte Verfahren, insbesondere aufgrund ihrer Geschwindigkeit, Anpassungsfähigkeit und einfachen Integration in digitale Lernumgebungen, zunehmend an Bedeutung für die akademische MCQ-Erstellung (Abb. 1).

Frühere Arbeiten zeigen, dass digitale Lehrmethoden – insbesondere im Bereich der medizinischen Biochemie – Kreativität fördern und die Motivation der Studierenden steigern können (M. Darabi et al., 2013; M. Darabi & Darabi, 2011). Aufgrund ihrer konzeptionellen Komplexität und der engen Verknüpfung mit der klinischen Praxis stellt die medizinische Biochemie einen besonders anspruchsvollen Prüfstein für die didaktische Validität KI-generierter MCQs dar. Die vorliegende Studie geht daher der Frage nach, inwieweit KI-generierte MCQs den didaktischen Anforderungen der medizinisch-biochemischen Ausbildung gerecht werden können.

Abbildung 1: Workflow der promptbasierten (»Closed-Book«) MCQ-Generierung mit einem LLM: Prompt → MCQ-Entwurf → Expert:innenbewertung vor dem Einsatz in der Lehre.



Methodik

Diese Evaluationsstudie verwendete ein vergleichendes Design (siehe Abb. 2), um die Leistungsfähigkeit von drei webbasierten KI-Plattformen zu bewerten: Gemini 2.5 Pro, Microsoft Copilot (GPT-4 Turbo) und ChatGPT-4. Mithilfe eines standardisierten Prompts generierte jede Plattform 20 MCQs zu drei zentralen Themenbereichen der medizinischen Biochemie: (i) biochemische Signalwege, (ii) Stoffwechselregulation und (iii) krankheitsbezogene Mechanismen. Die Fragen wurden so konzipiert, dass sie, gemäß der Bloom'schen Taxonomie, unterschiedliche kognitive Niveaus abbilden – von Reproduktion und Verständnis bis hin zu Anwendung und Analyse. Anschließend wurden die Fragen von Expert:innen entsprechend diesen vier Niveaustufen kategorisiert.

> **Prompt (verbatim):** Generate multiple-choice questions (MCQs) in medical biochemistry for undergraduate medical students. Create: 5 MCQs on biochemical pathways; 8 MCQs on metabolic regulation; 7 MCQs on disease-related mechanisms. Instructions: Each question must have one correct answer and three plausible distractors (a, b, c, or d). Ensure a mix of cognitive levels: ~30 % recall-based; ~30 % understanding; ~20 % application; ~20 % analysis.

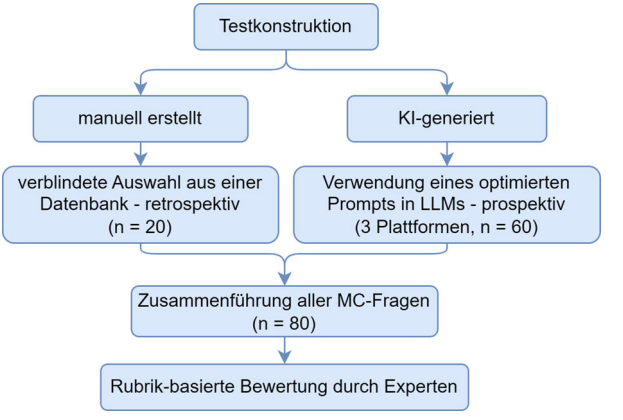
Insgesamt wurden demnach 60 KI-generierte MCQs mit 20 von Dozent:innen erstellten Fragen verglichen. Die menschlich erstellten Fragen wurden zufällig aus einem validierten institutionellen Fragenpool ausgewählt, um inhaltliche Relevanz und Diversität sicherzustellen. Fünf unabhängige Expert:innen aus der Biochemie

und medizinischen Didaktik führten, ohne vorherige Kenntnis über die Herkunft der jeweiligen Fragen, eine verblindete Bewertung durch.

Ein siebenkriterielles Bewertungsraster wurde verwendet: (1) Korrektheit der richtigen Antwort, (2) Plausibilität der Distraktoren, (3) Verständlichkeit der Fragestellung, (4) Item-Schwierigkeit (zielgruppenbezogen; unabhängig von Bloom), (5) sprachliche Qualität, (6) Lernzielrelevanz (Modulhandbuch) und (7) Vermeidung von Hinweisen (keine Cues im Fragetext/Antwortoptionen).

Gruppenunterschiede wurden mit dem Kruskal–Wallis-Test geprüft; Paarvergleiche erfolgten mittels Mann–Whitney-U-Test (Holm-korrigiert).

Abbildung 2: Schematische Übersicht des Studiendesigns. Bewertet wurden 80 MCQs (je 20 KI-generierte aus drei webbasierten LLM-Plattformen sowie 20 von Dozent:innen erstellte). Fünf Fachexpert:innen beurteilten alle Fragen anhand eines siebenkriteriellen Rasters; zudem wurden sie einer von vier Bloom-Stufen zugeordnet.



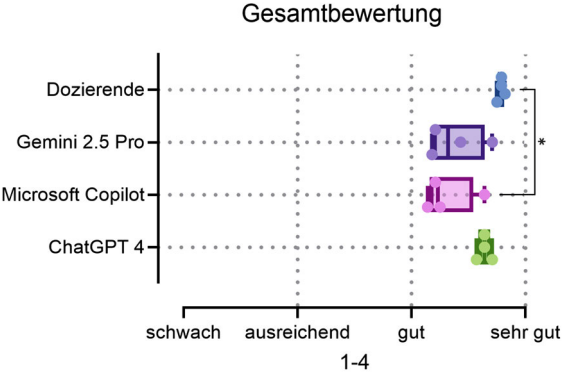
Ergebnisse

Insgesamt zeigten alle KI-Plattformen eine zufriedenstellende Leistung. Die von Dozent:innen erstellten MCQs erhielten, innerhalb der verblindeten Auswertung, die beste Gesamtbewertung, wobei die Punktverteilung deutlich in Richtung der Kategorie »Sehr gut« tendierte. Unter den KI-Plattformen kam ChatGPT-4 der Qualität menschlich generierter MCQs am nächsten.

Gemini 2.5 Pro und Microsoft Copilot erzielten ebenfalls Ergebnisse im Bewertungsbereich »Gut«, zeigten jedoch eine größere Variabilität über die Bewertungskriterien hinweg, was auf eine weniger konsistente Output-Qualität hinweist. Zwi-

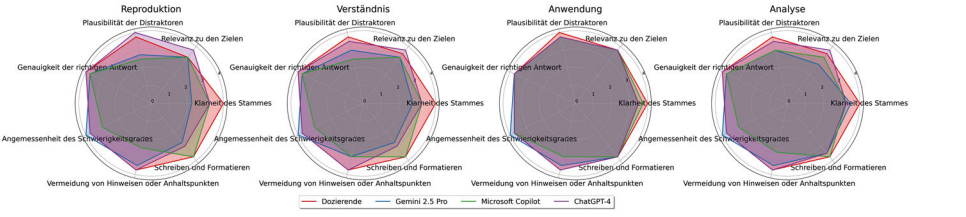
schen den menschlich erstellten Fragen und den von Gemini 2.5 Pro generierten MCQs zeigte sich ein statistisch signifikanter Unterschied ($p < 0{,}05$; siehe Abb. 3).

Abbildung 3: Vergleich der Gesamtbewertung von MCQs, erstellt von menschlichen Autor:innen und KI-Plattformen. Die Bewertungen erfolgten durch Expert:innen anhand einer vierstufigen ordinalen Skala (1 = schwach, 2 = ausreichend, 3 = gut, 4 = sehr gut). Boxplots zeigen Median und Interquartilsbereich; Datenpunkte entsprechen den Mittelwerten der fünf Expert:innenbewertungen pro kognitivem Niveau.



Die Analyse des kognitiven Anspruchsniveaus sowie der sieben Bewertungskriterien zeigte, dass ChatGPT insgesamt die beste Leistung erbrachte – insbesondere hinsichtlich der Korrektheit der Antworten und der Plausibilität der Distraktoren. Gemini 2.5 Pro erzielte schlechtere Bewertungen in Bezug auf die Verständlichkeit der Fragestellung und die Plausibilität der Distraktoren, während Microsoft Copilot die größten Schwierigkeiten bei der Generierung plausibler Distraktoren aufwies (siehe Abb. 4).

Abbildung 4: Vergleich der Leistung von drei KI-Plattformen und manuell erstellten Dozierendenfragen über vier Bloom-Stufen. Bewertet wurden sieben Bewertungskriterien: Korrektheit, Distraktor-Plausibilität, Verständlichkeit, (zielgruppenbezogener) Schwierigkeitsgrad, sprachliche Qualität, Lernzielrelevanz und Vermeidung von Hinweisen. AdS = Angemessenheit des Schwierigkeitsgrads; GdrA = Genauigkeit der richtigen Antwort; KdS = Klarheit des Stamms; PdD = Plausibilität der Distraktoren; RzdZ = Relevanz zu den Lernzielen; SuF = Sprache und Formatierung; VvHoA = Vermeidung von Hinweisen oder Anhaltspunkten.



Auf den niedrigeren kognitiven Ebenen wie Reproduktion und Verständnis schnitten sowohl Gemini als auch Copilot schlechter ab als die menschlich erstellten Fragen und ChatGPT. Besonders bei Gemini beeinträchtigten unklare Fragestellungen die Verständlichkeit. Bei Fragen auf der Bloom-Stufe Analyse zeigten Gemini und Copilot weiterhin Schwächen hinsichtlich der Plausibilität der Distraktoren und der Vermeidung unbeabsichtigter Hinweise, während ChatGPT und die Dozierendenfragen hier ausgewogenere Ergebnisse lieferten. Im Bereich der Anwendung war die Leistung über alle KI-Plattformen hinweg relativ konsistent, mit Ergebnissen, die sich der Qualität menschlich erstellter MCQs annäherten. Gemini erzeugte tendenziell klinisch reichhaltigere Szenarien, während die Stärke von Copilot in der Generierung strukturierter Vorlagen lag (siehe Abb. 4).

Diskussion

Die vorliegende Studie zeigt, dass LLM-basierte KI-Plattformen – Gemini 2.5 Pro, Microsoft Copilot und ChatGPT-4 – MCQs von angemessener Qualität im Bereich der medizinischen Biochemie generieren können, wobei sich je Plattform unter dem verwendeten standardisierten Prompt unterschiedliche Stärken im Output zeigten. Gemini 2.5 Pro lieferte klinisch reichhaltige Fragestellungen, Copilot zeigte einen Schwerpunkt auf Fragen zu biochemischen Signalwegen, häufig in vorstrukturierter Form, und ChatGPT-4 integrierte themenübergreifende Inhalte in prägnanter und kohärenter Formulierung. Auf höheren kognitiven Ebenen entsprachen die KI-generierten Fragen häufig der Qualität menschlich erstellter MCQs oder übertrafen diese sogar, was wiederum ein besonderes Potenzial für den Einsatz in der Bewertung der kognitiven Domänen Anwendung und Analyse erken-

nen lässt. Zu den häufigen Schwächen zählten jedoch unklare oder unzureichend spezifizierte Fragestellungen, Distraktoren, die redundant (inhaltlich sehr ähnlich) oder im Schwierigkeitsniveau unpassend waren, sowie gelegentliche Defizite in der klinischen Relevanz. Insgesamt kann prompt-basierte KI eine wertvolle Ergänzung zu traditionellen Fragenpools darstellen, bedarf aber zum aktuellen Zeitpunkt noch der wissenschaftlichen Einordnung und fachlichen Überprüfung von Expert:innen. Die unterschiedlichen Stärken der einzelnen Plattformen verdeutlichen die Relevanz einer kontextbezogenen Auswahl geeigneter KI-Tools in Abhängigkeit vom Bildungsziel und kognitivem Anspruchsniveau. Es sollte darauf hingewiesen werden, dass die Outputqualität nicht nur vom jeweiligen System, sondern auch von der Promptgestaltung und der plattformspezifischen Verarbeitung abhängt; der hier verwendete standardisierte Prompt könnte daher die beobachteten Unterschiede beeinflusst haben. In Zukunft sind weitere Studien erforderlich, um die Promptsensitivität plattformübergreifend zu quantifizieren.

Praktische Implikationen

Die Ergebnisse legen nahe, dass KI die Vielfalt der Aufgabenpools in der medizinischen Ausbildung erweitern kann. Jede Plattform zeigte spezifische Stärken: Gemini erwies sich als besonders effektiv bei der Generierung klinisch reichhaltiger, fallbasierter Fragestellungen; ChatGPT überzeugte durch prägnante und integrative Fragen zu verschiedenen Themenbereichen; und Microsoft Copilot bot strukturierte Vorlagen, die als nützliche Ausgangspunkte für die wissenschaftliche Praxis dienten. Diese Unterschiede unterstreichen die Notwendigkeit von Transparenz, wenn KI zur Prüfungsentwicklung beiträgt: Studierende sollten über den Einsatz informiert werden, und die verwendeten Prompts sowie zentrale Rahmenbedingungen (z.B. Modell/Version, Datum, Einstellungen) sollten dokumentiert werden, um auch zukünftig methodische Nachvollziehbarkeit und reproduzierbare Qualität zu gewährleisten.

Fazit und Ausblick

KI-generierte MCQs können die medizinische Ausbildung sinnvoll unterstützen, sofern sie sorgfältig integriert werden. Prompt-basierte Ansätze zeigen vielversprechendes Potenzial für den direkten Einsatz in Lehre und Prüfung, häufig mit nur geringem Nachbearbeitungsaufwand. Dennoch sind Expert:innen-basierte Überprüfungen entscheidend, um fachliche Bildungsstandards zu wahren und Verzerrungen vorzubeugen.

Zukünftige Forschung sollte untersuchen, wie sich KI-generierte Prüfungsfragen auf Lernergebnisse auswirken, wenn sie in realen Prüfungssituationen eingesetzt werden. Zudem sollten Strategien zur Minimierung von Bias entwickelt und evaluiert werden. Die Kombination komplementärer Stärken verschiedener KI-Plattformen könnte robuste hybride Workflows ermöglichen, die sowohl die Effizienz als auch die Qualität medizinischer Prüfungsformate verbessern.

Literatur

Abd-alrazaq, Alaa/AlSaad, Rawan/Alhuwail, Dari/Ahmed, Ahmed/Healy, Padraig M./Latifi, Syed/Aziz, Sarah/Damseh, Rafat/Alabed Alrazak, Sadam/Sheikh, Javaid (2023): »Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions«, in: JMIR Medical Education, 9, e48291.

Darabi, M./Darabi, M. (2011): Study on electronic designing and submission for Clinical Biochemistry course assignments as a teaching method, in: Tabriz University of Medical Sciences, 4th National Conference of E-Learning in Medical Sciences.

Darabi, M./Yazdi, Z./Darabi, M./Fayezi, S./Askari, M./Sarchami, R. (2013): »Infrastructure and Faculty Member Readiness for E-learning Implementation: The Case of Qazvin University of Medical Sciences«, in: Iranian Journal of Medical Education, 13(9), S. 730–740. https://ijme.mui.ac.ir/article-1-2695-fa.html.

Grévisse, Christian/Pavlou, Maria A. S./Schneider, Jochen G. (2024): »Docimological Quality Analysis of LLM-Generated Multiple Choice Questions in Computer Science and Medicine«, in: SN Computer Science, 5(5), S. 636.

Hang, Ching N./Wei Tan, Chee/Yu, Pei-Duo (2024): »MCQGen: A Large Language Model-Driven MCQ Generator for Personalized Learning«, in: IEEE Access, 12, S. 102261–102273.

Kunuku, Mourya T./Dehbozorgi, Narsin (2025): »Exploring Multimodal Quiz Generation and Evaluation Aligned with Higher-Order Learning Objectives in Bloom's Taxonomy«. in: Alexandra I. Cristea/Erin Walker/Yu Lu/Olga C Santos/Isotani, Seji, Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium, Blue Sky, and WideAIED, 2590, S. 433–438.

Learning Analytics

Integration von Learning Analytics in der Hochschullehre mittels POLARIS

Erfahrungen aus dem Praxisprojekt ISLAnD_BWL[1]

Holger Ketteniß und Johanna Marie Kunsmann

Individualisierte Lehr- und Lernangebote können Studierende bei der Bewältigung von Herausforderungen im Lernprozess unterstützen. Das Projekt ISLAnD_BWL entwickelt hierfür ein integriertes Learning Analytics System zur Erstellung individueller Development-Pfade und adressiert zunächst ein BWL-Pflichtmodul, welches sich durch eine Vielzahl digitaler Lehrmaterialien auf verschiedenen Plattformen auszeichnet. Bei Studierenden des Moduls herrscht oftmals Unsicherheit darüber, ob die ausgewählten Lehrmaterialien zur Erreichung des erforderlichen Kompetenzniveaus ausreichen. Durch den pilotierten Anschluss des elektronischen Prüfungs- und Übungssystems »Dynexite« an das durch KI:edu:NRW entwickelte POLARIS-System wird der Lernfortschritt der Studierenden auf Basis der erbrachten Leistungen analysiert. Dies ermöglicht individuelle Vorschläge zur Nutzung der Lernmaterialien, sodass Studierenden durch individuelle Lernsettings beim erfolgreichen Bestehen des Moduls unterstützt werden.

Integration of learning analytics in higher education teaching with POLARIS – lessons learned from the ISLAnD_BWL project

Individualized teaching and learning offerings can support students in overcoming challenges within their learning process. The project ISLAnD_BWL is developing an integrated learning analytics system for creating individual development paths and is initially addressing a mandatory business administration module (›BWL‹), which is characterized by a wide range of digital teaching materials on various platforms. Students of the module often feel unsure whether the selected teaching materials will enable them to achieve the required level of competence. Through the piloted connection of the electronic examination and exercise system ›Dynexite‹ to the POLARIS system funded by KI:edu:NRW, students learning progress is analyzed on the basis of their continuing performance. This allows for individual suggestions on the use of learning materials, so that students are supported in successfully completing the module through individual learning settings.

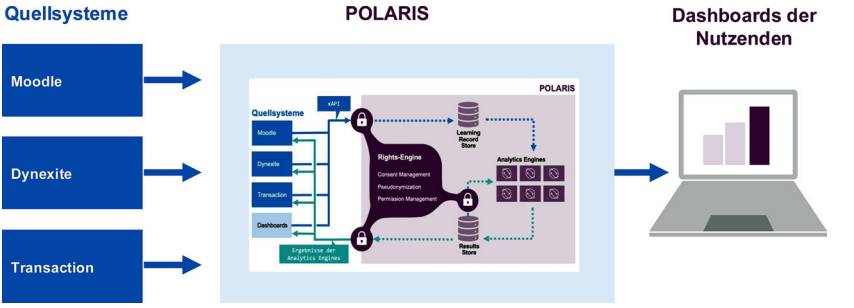
1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

Einleitung und Projektvorstellung

Zahlreiche Studierende sind im Verlauf ihres Studiums mit Herausforderungen konfrontiert, die ihren Lernfortschritt und damit auch ihren Studienerfolg maßgeblich beeinflussen. Ein zentrales Problem ist das Zeitmanagement, da es den Studierenden oftmals schwerfällt, die ihnen zur Verfügung stehende Zeit effizient auf die einzelnen zu absolvierenden Module zu verteilen. Hinzu kommt die Unsicherheit, ob die ausgewählten Lern- und Lehrmaterialien ausreichend sind, um das notwendige Kompetenzniveau für das Bestehen eines Moduls zu erreichen (Fook/Sidhu 2015: 608–610). Denn oftmals stellen Lehrende eine Vielzahl an Lern- und Lehrmaterialien zur Verfügung, um den heterogenen Bedürfnissen der Studierenden gerecht zu werden. Dies kann jedoch dazu führen, dass die Studierenden mit der strategischen Auswahl der für sie jeweils geeigneten Materialien überfordert sind. Insbesondere in den ersten Fachsemestern sowie bei Studierenden mit bislang schwächeren akademischen Leistungen stellt die selbstständige Entwicklung einer passenden individuellen Lernstrategie eine große Herausforderung dar (García-Pérez/Fraile/Panadero 2021: 544–546). Hinzu kommt das Problem, dass Studierende und auch Lehrende Lernprobleme oftmals erst kurz vor der Klausur bzw. anhand der Klausurergebnisse am Ende des Semesters erkennen. Diese verspätete Erkenntnis verhindert wirksame, semesterbegleitende Maßnahmen, sodass eine frühzeitige Intervention und zielgerichtete Unterstützung der Studierenden während des Semesterverlaufs nicht möglich sind (Larrabee Sønderlund/Hughes/Smith 2019: 2614–2616). Eine Lösung zur Bewältigung dieser Herausforderungen kann die Etablierung eines Learning Analytics Systems sein. Ein solches ermöglicht Studierenden individualisierte Lehr- und Lernangebote sowie kontinuierliche Feedbackschleifen und bietet Lehrenden einen Überblick über den Lernfortschritt der Studierenden (Schumacher/Ifenthaler 2018: 612–615).

Das Projekt ISLAnD_BWL entwickelt hierfür ein integriertes Learning Analytics System zur Erstellung individueller Lern- und Entwicklungs-Pfade (»Development«). Im Rahmen eines BWL-Pflichtmoduls, erfolgt ein pilotierter sukzessiver Anschluss der Lehrmaterialien und Lernplattformen (Quellsysteme) an das durch KI:edu.nrw entwickelte POLARIS-System (Platte/Thelen/Breuer 2025: 115). Ziel ist die Entwicklung eines quellsystemübergreifenden Dashboards, das Informationen zum Lernfortschritt der Studierenden auf unterschiedlichen Aggregationsebenen für verschiedene Nutzergruppen bereitstellt (siehe Abbildung 1).

Abbildung 1: Anbindung des Learning Analytics System »POLARIS«



Der aktuelle Entwicklungsstand des POLARIS-Systems ermöglicht bereits die Anbindung an die Lernplattform »Dynexite«, die ein elektronisches Prüfungs- und Übungssystem darstellt. Perspektivisch ist auch eine Anbindung an andere Lernplattformen (z.B. Moodle, Transaction) geplant. Die fortlaufende Analyse des Lernfortschrittes von Studierenden bietet für verschiedene Nutzergruppen wertvolle Einblicke (Romero/Ventura 2013: 25):

1. Für **Studierende** ermöglicht die Entwicklung individueller Lernsettings ein frühzeitiges Erkennen von Lernerfolgen und Lerndefiziten und unterstützt sie demnach beim Erreichen des notwendigen Ziel-Kompetenzniveaus des Moduls.
2. **Lehrende** erhalten Einblicke in den Lernfortschritt der Studierenden während des Semesters, wodurch sowohl mögliche Anpassungen der eingesetzten Lern- und Lehrmaterialien angeregt als auch Impulse für eine zielgerichtete didaktische Begleitforschung gegeben werden können.
3. Für das **Studienmanagement** ermöglicht insbesondere die perspektivisch geplante fächerübergreifende Anwendung des POLARIS-Systems einen erweiterten Einblick in den Studienverlauf der Studierenden. Dadurch ist eine gezielte Weiterentwicklung von Mentoring-Angeboten möglich.

Vorstellung des Pilotmoduls

Das Modul »Buchführung und internes Rechnungswesen«, welches eine Einführung in das betriebswirtschaftliche Rechnungswesen darstellt, wurde als Pilotierung für die Anbindung an das POLARIS-System an der RWTH Aachen ausgewählt. Das Modul ist im Curriculum zahlreicher Studiengänge – zumeist in den ersten Fachsemestern – verankert (z.B., BWL B.Sc., Informatik B.Sc., Wirtschaftsingenieurwesen B.Sc., Mathematik B.Sc., Digitale Medienkommunikation

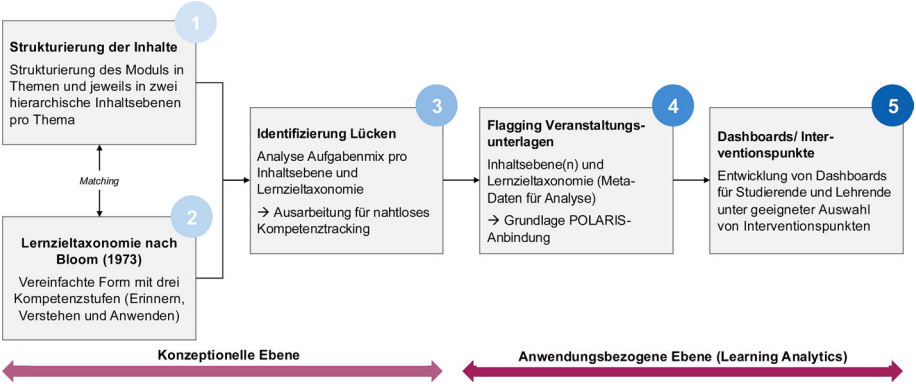
M.A.), sodass sich eine heterogene Studierendengruppe von ca. 1900 Modulteilnehmer*innen pro Semester ergibt. In den vergangenen Semestern wurden sowohl eine abnehmende Anwesenheit in den Präsenzveranstaltungen als auch eine steigende Durchfallquote festgestellt. Als erste Maßnahme wurde im Wintersemester 23/24 eine Prüfungszulassungsvoraussetzung eingeführt. Studierende müssen seitdem für die Zulassung zur Klausur eine Mindestanzahl an Online-Hausaufgaben in »Dynexite« erfolgreich absolvieren, die innerhalb eines einwöchigen Bearbeitungszeitraums flexibel von zuhause aus bearbeitet werden können und aufgeteilt auf sieben Online-Hausaufgaben alle relevanten Inhalte des Moduls abfragen. Das Projekt ISLAnD_BWL ist zentraler Bestandteil einer weiteren Maßnahme, um vor allem den Kompetenzerwerb unterhalb des Semesters weiter zu fördern und so die Studierenden beim erfolgreichen Abschluss des Moduls zu unterstützen.

Innerhalb des Pilotmoduls wird den Studierenden eine breite Auswahl an Lern- und Lehrmaterialien zur Verfügung gestellt. Das Modul wird in einem Flipped-Classroom-Format durchgeführt und es kommen verschiedene Lernplattformen zur Unterstützung des Blended Learnings zum Einsatz. Neben der Bereitstellung von Lehrvideos und Folienskripten handelt es sich insbesondere um den Einsatz des Serious Games »Transaction«, des »Dynexite« Systems, welches bereits zahlreiche parametrisierte Übungs- und Prüfungsaufgaben zu allen Inhalten des Moduls umfasst, sowie diverser Moodle-Funktionalitäten (z.B. H5P in Lehrvideos, Fortschrittslisten und Glossare). Zusätzlich kommen in den Präsenzsitzungen Audience-Response-Systeme zum Einsatz, um z.B. Live-Quizzes durchzuführen. Die Prüfungszulassung sowie die Prüfung werden in »Dynexite« absolviert.

Prozess zur Etablierung eines Learning Analytics Systems

Im Rahmen des Projektes wurde in Zusammenarbeit mit den Projektpartnern zunächst ein allgemein anwendbarer Prozess zur Integration eines Learning Analytics Systems in bestehenden Modulen entwickelt. Dieser Prozess ermöglicht eine systematische und effiziente Erweiterung der Anbindung des POLARIS-Systems auf weitere Module. Wie in Abbildung 2 dargestellt, kann zwischen einer konzeptionellen und anwendungsbezogenen Ebene unterschieden werden. Die Beschreibung der fünf Schritte wird im Folgenden jeweils durch Beispiele aus der Pilotierung veranschaulicht.

Abbildung 2: Prozessschritte zur Etablierung eines Learning Analytics Systems

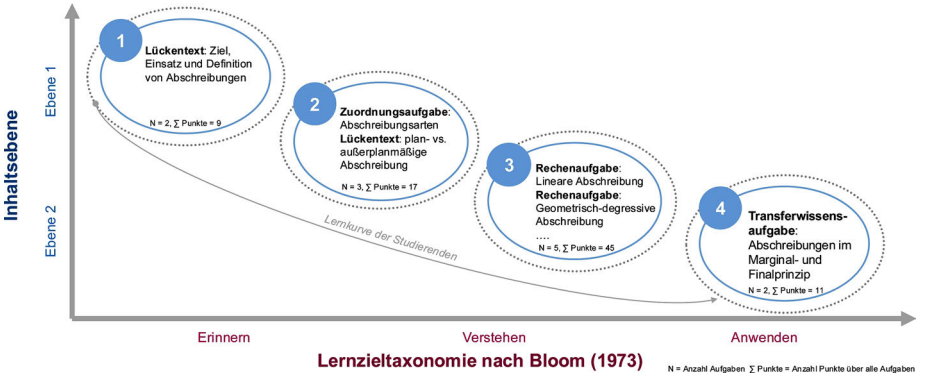


Im **ersten Schritt** ist eine Strukturierung der Inhalte des Moduls notwendig. Pro identifiziertem Themenbereich erfolgt zunächst eine Strukturierung der Inhalte in zwei hierarchische Inhaltsebenen. Im Anschluss findet eine Zuordnung aller Übungsaufgaben zu den jeweiligen Inhaltsebenen statt. Das Pilotmodul ist in die Teile »Buchführung« und »Internes Rechnungswesen« unterteilt. Der Teil »Buchführung« umfasst insgesamt drei Themen, während der Teil »Internes Rechnungswesen« vier Themen enthält. Jedes Thema lässt sich in weitere Unterebenen der ersten Hierarchieebene (»Inhaltsebene 1«) unterteilen, die aus jeweils mehreren Unterebenen der zweiten Hierarchieebene (»Inhaltsebene 2«) bestehen.

Innerhalb des **zweiten Schrittes** erfolgt eine erweiterte Zuordnung der Übungsaufgaben durch ein Matching dieser zu den Kompetenzstufen »Erinnern«, »Verstehen« und »Anwenden«, die in Anlehnung an die Lernzieltaxonomie nach Bloom et al. (1956: 18) definiert sind. Demzufolge werden nur die ersten drei Taxonomiestufen abgebildet, da in dem Pilotmodul einführende bzw. grundlegende Inhalte behandelt werden und ein Kompetenzerwerb auf den letzten drei Bloom'schen Taxonomiestufen nicht im Fokus steht. Im Zuge des beschriebenen Matchings werden Übungsaufgaben, bei denen ausschließlich das Erinnern von im Modul behandelten Definitionen und Begriffen gefordert ist, der Kompetenzstufe »Erinnern« zugeordnet. Übungsaufgaben, die die Wiedergabe der in dem Modul behandelten Inhalte auf einer einzelnen Inhaltsebene erfordern, werden der Stufe »Verstehen« zugeordnet. Übungsaufgaben, die mehrere thematisierte Inhalte auf verschiedenen Inhaltsebenen verknüpfen und damit eine Anwendung der thematisierten Inhalte erfordern, werden der Kompetenzstufe »Anwenden« zugeordnet. Letztlich ermöglicht das Matching der Übungsaufgaben zu den jeweiligen Inhaltsebenen und der Lernzieltaxonomie einen transparenten Überblick über den Mix der bislang verfügbaren Übungsaufgaben. Abbildung 3 stellt dies exemplarisch für die bisher verfügbaren Übungsaufgaben zur Inhaltsebene 1 Abschreibungen« sowie den dazugehö-

rigen Inhaltsebenen 2 (z.B. »lineares Verteilungsverfahren«, »geometrisch-degressives Verteilungsverfahren«) aus dem dritten Thema des Teils »Buchführung« der Pilotierung dar.

Abbildung 3: Beispielhafte Visualisierung des Schrittes 1 und 2



Im **dritten Schritt** erfolgt auf Basis einer Analyse der Aufgaben hinsichtlich der Anzahl und Gesamtpunktzahl pro Inhaltsebene sowie der Lernzieltaxonomie für die jeweiligen Themen des Moduls eine Identifizierung von Lücken in den Übungsaufgaben. Ein ggf. identifizierter Überarbeitsbedarf der Übungsaufgaben muss innerhalb des Schrittes umgesetzt werden, um aus didaktischer Perspektive eine stufenweise Kompetenzentwicklung der Lernenden zu fördern. Zudem ist dies für ein nahtloses Kompetenztracking durch die POLARIS-Anbindung notwendig. Das in Abbildung 3 dargestellte Beispiel zeigt, dass auf der Inhaltsebene 1 und der Lernzieltaxonomie »Erinnern« sowohl die Anzahl (N) als auch die Gesamtpunktzahl der Aufgaben (Σ) vergleichsweise gering ist. Dies gilt auch für die Inhaltsebene 2 und die Lernzieltaxonomie »Anwenden«.

Im **vierten Schritt** folgt die anwendungsbezogene Ebene, d.h. dort beginnen die systemseitigen Vorbereitungen für die Anbindung an das Learning Analytics System. Zunächst erfolgt ein standardisiertes Flagging der Übungsaufgaben, bei dem diesen sowohl die jeweils adressierten Inhaltsebenen als auch die dazugehörige Lernzieltaxonomie zugeordnet werden. Dies ermöglicht es dem POLARIS-System, auf Grundlage dieser erzeugten Meta-Daten eine differenzierte Analyse des Lernfortschrittes der Studierenden durchzuführen.

Anschließend erfolgt im **fünften Schritt** die Entwicklung der auf das Modul angepassten Dashboards für Studierende und Lehrende, welche den Lernfortschritt der Teilnehmer auf unterschiedlichen Aggregationsebenen visualisieren. Durch eine gezielte Auswahl von allgemeinen und inhaltlichen Interventionspunkten

erhalten Studierende individuelle Lernempfehlungen entsprechend ihres Lernfortschrittes, die visuelle Hinweise mit einer Auswahl geeigneter Lehrmaterialien verbinden (Wong/Li 2018: 181). Zudem wird Lehrenden ermöglicht frühzeitige didaktische Maßnahmen im Semesterverlauf zu ergreifen.

Fazit und Ausblick

Zusammengefasst erfordert die effiziente Implementierung eines integrierten Systems für Learning Analytics & Development Systems auf konzeptioneller Ebene eine Strukturierung und ggf. Überarbeitung der Lehrmaterialien (Zhong 2023: 7068). Auf anwendungsbezogener Ebene umfasst dies neben der technischen Umsetzung auch die Entwicklung sowohl allgemeiner als auch modulbezogener inhaltlicher Interventionspunkte.

Die entwickelten Prozessschritte zur Etablierung eines Learning Analytics Systems ermöglichen eine Übertragbarkeit auf andere Module. Auf konzeptioneller Ebene bieten sie Lehrenden einen Leitfaden zur zielgerichteten Weiterentwicklung von Lern- und Lehrmaterialien sowie zur didaktischen Nachbesserung der Kompetenzorientierung. Auf anwendungsbezogener Ebene ermöglichen die gewonnenen Daten Einblicke in den Lernfortschritt der Studierenden, woraus sich semesterbegleitenddidaktische Maßnahmen ableiten lassen. Für Studierende bietet die strukturierte und transparente Übersicht ihres Lernfortschrittes und der adressierten Kompetenzen des Moduls eine Orientierung im Lernprozess. Auch für das Studienmanagement ergibt sich ein zusätzlicher Erkenntnisgewinn. Insgesamt liefert das Praxisprojekt einen wertvollen Einblick in die Verzahnung von kompetenzorientierter Lehre und der Integration von Learning Analytics Systemen.

Literatur

Bloom, B. S./Engelhart, M. D./Frust, E. J./Hill, W. H./Krathwohl, D. R. (1956): Taxonomy of educational objectives – The classification of educational goals, in: New York: David McKay, S. 483–498.

Fook, Chan Yuen/Sidhu, Gurnam Kaur (2015): Investigating Learning Challenges Faced by Students in Higher Education, in: Procedia – Social and Behavioral Sciences 186, S. 604–612.

García-Pérez, Daniel/Fraile, Juan/Panadero, Ernesto (2021): Learning strat-egies and self-regulation in context: how higher education students ap-proach different courses, assessments, and challenges, in: Eur J Psychol Educ 36 (2), S. 533–550.

Larrabee Sønderlund, Anders/Hughes, Emily/Smith, Joanne (2019): The efficacy of learning analytics interventions in higher education: A sys-tematic review, in: Brit J Educational Tech 50 (5), S. 2594–2618.

Platte, Laura/Thelen, Kira/Breuer, Martin (2025): »Informierte Einwilligung für Learning Analytics. Gestaltung eines Informations- und Entschei-dungsdialogs«, 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, S. 111–118.

Romero, Cristobal/Ventura, Sebastian (2013): »Data mining in education«, in: WIREs Data Min & Knowl 3 (1), S. 12–27.

Schumacher, Clara/Ifenthaler, Dirk (2018): »The importance of students' motivational dispositions for designing learning analytics«, in: J Comput High Educ 30 (3), S. 599–619.

Wong, Billy T.M./Li, Kam Cheong (2018): »Learning Analytics Interven-tion: A Review of Case Studies«, in: 2018 International Symposium on Educational Technology (ISET). 2018 International Symposium on Educational Technology (ISET). Osaka, Japan, 31.07.2018 – 02.08.2018: IEEE, S. 178–182.

Zhong, Lin (2023): »A systematic review of personalized learning in higher education: learning content structure, learning materials sequence, and learning readiness support«. in: Interactive Learning Environments 31 (10), S. 7053–7073.

AnnoPy – Potenziale von KI-basierter Learning Analytics für die Förderung wissenschaftlicher Textkompetenz[1]

Oliver Scholle, Sara Rezat, Sebastian Rezat und Dominik Niehus

AnnoPy ist ein digitales Werkzeug, um kollaborativ Texte zu lesen, zu erschließen und zu analysieren. Es ist für die Präsenzlehre konzipiert und unterstützt Lehrende dabei, den Übergang von der individuellen Auseinandersetzung mit wissenschaftlichen Texten hin zu einer kollaborativ-sozialen Aushandlung des Gelesenen in Lehrveranstaltungen zu organisieren und zu strukturieren. Im Beitrag wird das digitale kollaborative Tool AnnoPy vorgestellt und gezeigt, welche Potenziale sich für die Analyse von Daten über Lernende durch die Integration von KI-basierter Learning Analytics in das Tool ergeben. Anhand von exemplarischen Daten aus einer Vorlesung der Mathematikdidaktik wird zunächst vorgestellt, wie in AnnoPy eine algorithmisch-händische Analyse und Datenauswertung erfolgte, um dann diese Art der Auswertung mit einer KI-gestützten Auswertung, bei der ein MCP-Server eingebunden wurde, zu vergleichen.

AnnoPy – Potentials of AI-based learning analytics for promoting academic writing skills

AnnoPy is a digital tool for collaboratively reading, exploring and analysing texts. It is designed for classroom teaching and supports teachers in organising and structuring the transition from individual engagement with academic texts to collaborative social negotiation of what has been read in teaching sessions.

This paper introduces the digital collaborative tool AnnoPy and shows the potential for analysing and evaluating data about learners by integrating AI-based learning analytics into the tool. Using sample data from a mathematics lecture for prospective primary teachers, it first shows how AnnoPy performed an algorithmic-manual analysis and data evaluation and then compares this type of evaluation with an AI-supported evaluation that integrated an MCP server.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

Einleitung

In der universitären Lehre spielt der rezeptive und produktive Umgang mit Texten für den Erwerb grundlegenden fachlichen Wissens noch immer die zentrale Rolle. Wissenschaftliche Texte (Forschungsbeiträge, Lehrbücher, Skripte etc.) sind zentrale Ausgangs- und Bezugspunkte des fachlichen Lernens, denn Wissen wird nicht nur in Texten niedergelegt, sondern auf der Grundlage von Texten erworben (vgl. Feilke 2001). Für diesen fachlichen Wissenserwerb im Studium ist eine entsprechende Textkompetenz (vgl. Portmann-Tselikas/Schmölz-Eibinger 2008: 5–8.) eine notwendige Voraussetzung. Zum einen müssen Studierende über grundlegende Lese-, Schreib- und sprachreflexive Kompetenzen bereits beim Eintritt ins Studium verfügen, zum anderen gilt es, diese literalen Kompetenzen domänenspezifisch, d.h. bezogen auf die jeweiligen Anforderungen im Fach, im Laufe des Studiums auszubauen. Studierende müssen lernen, wissenschaftliche Fachtexte zu rezipieren und inhaltlich zu durchdringen. Gefordert wird aber auch das eigenständige Produzieren wissenschaftlicher Texte in Form von Hausarbeiten bzw. Abschlussarbeiten. Aus didaktischer Sicht ist hier die enge Verzahnung von Lesen und Schreiben wichtig, weil sich diese positiv auf die Entwicklung von Textkompetenzen auswirkt (vgl. entsprechende Metastudien u.a. Alves et al. 2020).

Die Entwicklung dieses Erwerbsprozesses, die Bedingungsfaktoren und die damit verbundenen Hürden sind mittlerweile gut durch entsprechende empirische Untersuchungen dokumentiert (vgl. u.a. Kruse 2010). Lehrende stehen im Rahmen universitärer Lehrveranstaltungen nun konkret vor der didaktischen Herausforderung, die individuelle Auseinandersetzung der Studierenden mit einem wissenschaftlichen Text einzuschätzen und dies bei der Unterstützung und Gestaltung des Lernprozesses in der gesamten Lerngruppe angemessen zu berücksichtigen. Übergreifendes Ziel ist dabei, dass Studierende die Fähigkeit erwerben, angemessen wissenschaftliche Texte zu rezipieren.

Einen Einblick in die Frage, wie der/die einzelne Studierende einen (Fach-)Text erschlossen hat und welche Probleme es dabei gab, erhalten Lehrende in klassischen Lehrveranstaltungen in der Regel nur vereinzelt auf Nachfrage, in Vorlesungen in der Regel gar nicht. Diese didaktische Herausforderung adressieren wir mit der Entwicklung und dem Einsatz des Tools *AnnoPy*.

Im Beitrag werden wir zunächst die Funktionalitäten des Tools *AnnoPy* (www.annopy.de) vorstellen. Um die Auswertung der in *AnnoPy* generierten Daten effizienter auswerten zu können und einen gezielten Einblick in die Daten der Lerngruppe zu erhalten, ist in einer neueren Version von *AnnoPy* eine KI-basierte Learning Analytics integriert worden. Damit ist es möglich, individuelle ebenso wie gruppenübergreifende Problembereiche beim Umgang mit Fachtexten bezogen auf den Lernprozess und das Lernergebnis zu identifizieren. Anhand eines Beispiels aus einer ma-

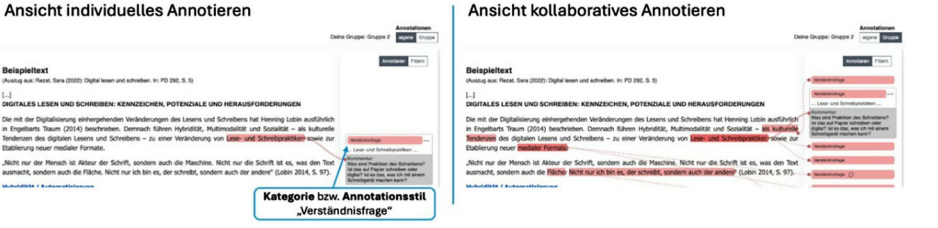
thematikdidaktischen Vorlesung wird erläutert, warum und wie die Integration der KI-basierten Learning Analytics erfolgte und welche Potenziale damit einhergehen.

Funktionalitäten des Tools *AnnoPy*

AnnoPy ist ein digitales Werkzeug, um kollaborativ Texte zu lesen, zu erschließen und zu analysieren. Es unterstützt Lehrkräfte dabei, den Übergang von der individuellen Auseinandersetzung mit Texten hin zu einer kollaborativ-sozialen Aushandlung des Gelesenen im Unterricht zu organisieren und automatisiert zu strukturieren. Lernenden bietet *AnnoPy* kognitiv aktivierende und sozial-interaktive Lernaktivitäten im Umgang mit Texten. Die Besonderheit von *AnnoPy* liegt darin, dass es im digitalen Raum innovative und inklusive Formen der Partizipation und des social readings ermöglicht, die in analogen Formaten oft schwer umzusetzen sind. Lernende können ihre individuellen Erkenntnisse aus der Auseinandersetzung mit Texten aktiv einbringen und diese im Zusammenspiel mit der gesamten Lerngruppe reflektieren. Dies fördert ein tiefgreifendes Verständnis – nicht nur auf inhaltlicher, sondern auch auf sprachlicher und struktureller Ebene von Texten – und adressiert damit die für die Entwicklung von Textkompetenz wichtige Verzahnung von Lesen und Schreiben.

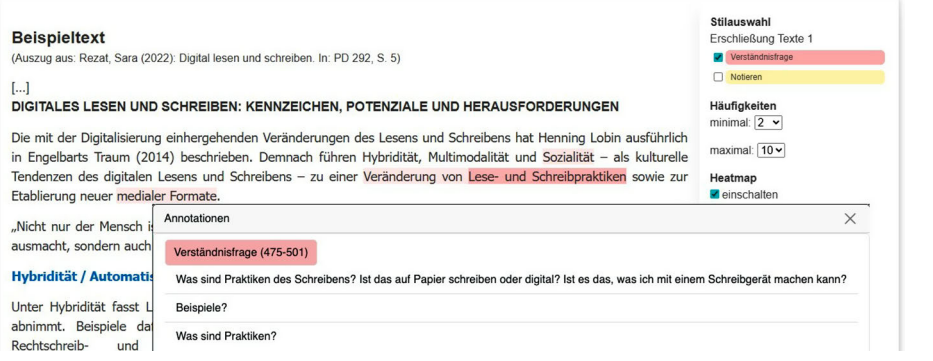
Der Arbeitsprozess mit *AnnoPy* vollzieht sich grundsätzlich in drei Phasen, die hier in einem prototypischen Ablauf vorgestellt werden. Zunächst setzen sich die Lernenden individuell mit einem Text auseinander und annotieren (markieren und ggf. kommentieren) ihn auf Grundlage relevanter Kriterien, z. B. welche Informationen für sie neu sind oder wo Verständnisfragen geklärt werden müssen (vgl. Abb. 1). Die Kriterien (sog. Annotationsstile) können von Lehrkräften in *AnnoPy* im Vorfeld flexibel – je nach Zielsetzung – definiert werden. In dieser Phase steht das problemlösende Handeln des Individuums im Vordergrund. Die Funktionalitäten des Tools erlauben es aber auch, die individuelle Lektüre um eine kollaborative Zwischenphase zu erweitern, bei der die einzelnen Lernenden nicht nur ihre Annotationen sehen und bearbeiten können, sondern auch die ihrer Gruppe (vgl. Abb. 1, rechts). Auf diese Weise ist ein Vergleich des eigenen Vorgehens mit den Gruppenmitgliedern möglich.

Abbildung 1 links: Ansicht individuelles Annotieren, rechts: Ansicht kollaboratives Annotieren



Die grundlegende Funktion des Tools besteht darin, dass das System in Echtzeit die individuellen Annotationen und Kommentare der gesamten Lerngruppe automatisch wie Folien übereinanderlegt. Es entstehen damit aggregierte und anonyme Daten der gesamten Lerngruppe. Dieses sogenannte Overlay macht alle individuellen Auseinandersetzungen mit dem Text kumuliert sichtbar und bildet die Grundlage für die Anschlusskommunikation bzw. für kollaborative Aushandlungsprozesse mit der gesamten Lerngruppe bzw. im Plenum. Hierfür erfolgt nach der individuellen Annotation eine algorithmische Aufbereitung für die Visualisierung und quantitative Analyse aller Annotationen und Kommentare, die die Grundlage für die dritte Phase bilden. Hier findet die von der Lehrkraft moderierte Anschlusskommunikation unter Verwendung der in der Auswertungsansicht bereitgestellten Filter- und Visualisierungswerkzeuge statt. Das Overlay der Annotationen ermöglicht einen shared focus im Sinne sozialkonstruktivistischer Konzeptionen und didaktische sinnvolle Fokussierungen von relevanten Textstellen (vgl. Behrens/Madlener-Charpentier 2022: 41). Wie dies mithilfe der Auswertungsansicht erfolgen kann, wird am folgenden Beispiel erläutert.

Abbildung 2: Auswertungsansicht mit aktivierter Annotation der Passage »Lese- und Schreibpraktiken«.



Mit den bereitgestellten Filterwerkzeugen kann die Lehrkraft (ggf. auch vorbereitend) auswählen, welche und wie viele annotierte und kommentierte Textstellen beim Overlay angezeigt werden. Bei der Auswertung aus Abb. 2 sind für das Sprechen über den Text Passagen ausgewählt, bei denen es Verständnisschwierigkeiten gibt. Des Weiteren ist die Heatmap-Funktion aktiviert, um die Stellen schneller identifizieren zu können, die am häufigsten markiert wurden. Interessant für eine produktive Anschlusskommunikation können z.B. in anderen Szenarien auch Textstellen sein, bei denen durch die abweichenden Annotationen ein Dissens offenbar wird. Die quantitative Auswertung der Annotationsdaten generiert auf diese Weise digitale Lernartefakte, die für die Anschlusskommunikation relevant sein können. Die Einblendung der Kommentare erfolgt in dieser Ansicht durch einen Klick/Touch auf die betreffenden Stellen.

Die Funktionsweise des Tools und die beschriebene Einsatzweise bilden die Grundlage für die Auseinandersetzung mit wissenschaftlichen Texten in allen Studiengängen bzw. Fächern. Es sind insgesamt drei didaktische Einsatzszenarien zu unterscheiden, die mit den Funktionalitäten des Tools umgesetzt werden können. Das erste Szenario, das Erschließen bzw. Verstehen von Texten – an dem zuvor die Funktionsweise des Tools erläutert wurde – fokussiert die intensive Auseinandersetzung mit Fachtexten. Weiterhin stellen die Textanalyse, bei der z.B. die sprachlichen Gestaltungsmittel wissenschaftlicher Texte im Fokus stehen (vgl. Kruse 2010: 149–182) und Feedback zu Texten weitere Szenarien zur Förderung wissenschaftlicher Textkompetenzen (vgl. Rezat et al. 2024: 222–223) dar. Alle Szenarien zielen darauf ab, rezeptive Textkompetenzen der Studierenden im Umgang mit wissenschaftlichen Texten zu fördern, die sich wiederum positiv auf die produktiven Textkompetenzen auswirken. Darüber hinaus ist *AnnoPy* auch ein Diagnosetool, weil durch die Bereitstellung des Overlays und der Auswertungswerkzeuge ein diagnostischer Einblick in die Lernergebnisse der Gruppe möglich wird. Auf diese Weise können zentrale Fragen, Problembereiche und Anmerkungen einer Lerngruppe bei der Vorbereitung einer Veranstaltung zur adaptiven Gestaltung berücksichtigt werden.

Grundlage für die beiden genannten Funktionen bilden die in der Datenbank gespeicherten Markierungen und Kommentare aller User am Text. Die Daten sind in diesem Sinne ein kollaboratives Produkt einer spezifischen Lerngruppe und die einzelnen Annotationen mit ihren Attributen, wie in Abb. 3 dargestellt, bilden die Datenbasis.

Abbildung 3: Auszug Datensatz Beispiel Annotation aus Abb. 2, links

pk	Inhalt	Kommentar	Stil	Start	Ende	Ersteller:in	Rolle	Gruppe	erstellt	aktualisiert
p9jk7oM	Lese- und Schreibpraktiken	Was sind Praktiken des Schreibens? Ist das auf Papier schreiben oder digital? Ist es das, was ich mit einem Schreibgerät machen kann?	Verständnisfrage	475	501	[redacted]	Teilnehmer:in	Gruppe 2	21.08.2025 14:13:22	21.08.2025 14:14:01

Probleme der quantitativen algorithmischen Auswertung der Daten in *AnnoPy*

Durch das kollaborative digitale Annotieren und Kommentieren von Texten mit *AnnoPy* können sehr schnell große Datenmengen entstehen. Die Anzahl der Lernenden, die vorgegebenen Stile oder die Länge des Textes können dafür sorgen, dass mehrere tausend Annotationen (und Kommentare) an einem Text vorgenommen werden und die vorgestellten Annotations- und Auswertungswerkzeuge an ihre Grenzen stoßen. Die Abb. 4 zeigt einen Ausschnitt aus einem Skript der Vorlesung »Elemente der Geometrie« (Didaktik der Mathematik, Universität Paderborn), das in *AnnoPy* von 270 Studierenden im Vorfeld der Vorlesung annotiert und kommentiert wurde.

Abbildung 4: links: Auswertungsansicht ohne Filterung, rechts: Auswertungsansicht mit Filtern (Heatmap, nur Stellen mit mind. 20 Annotationen)

Die Ansichten in Abb. 4 veranschaulichen die zuvor benannten Grenzen der bisherigen Auswertungs- und Visualisierungsmöglichkeiten. Auch die in Abb. 4: rechts mithilfe der Filter generierte Ansicht erschwert die Identifikation relevanter Textstellen oder Kommentare. Eine Nutzung der Ansicht ohne eine umfangreiche vorgeschaltete aufwendige Analyse solcher Daten ist nicht möglich. Stattdessen mussten zur Vorbereitung der Sitzung die Annotationsdaten händisch von einer wissenschaftlichen Hilfskraft unter Verwendung der Filterfunktionen zeitaufwändig ausgewertet werden. Vor allem für die Auswertung der Kommentare, die eine qualitativ-inhaltliche Analyse erfordert, war die Hilfskraft unerlässlich. Somit stand auf der einen Seite als positives Ergebnis des Tool-Einsatzes die Aktivierung der Studierenden, die sich intensiv mit dem Text befasst haben, und die zahlreichen Annota-

tionsdaten, die es ermöglichten, bei der Durchführung der Sitzung die Fragen und Anmerkungen der Studierenden einzubinden. Auf der anderen Seite mussten wir feststellen, dass die Auswertungswerkzeuge noch nicht hinreichend automatisiert sind, um sie didaktisch funktional und zielführend nutzen zu können.

Integration KI-basierter Learning Analytics in *AnnoPy*

Die Lösung der bestehenden Probleme bei der Auswertung und Aufbereitung der Daten bestand darin, eine KI-gestützte Learning Analytics in *AnnoPy* zu integrieren, mit dem Ziel, die Arbeit der Hilfskraft an einen KI-Agenten zu übertragen, der mithilfe bereitgestellter Werkzeuge und Daten für unterschiedliche Szenarien Auswertungen automatisiert durchführen kann.

Wenn wir im Folgenden von KI-gestützter Learning Analytics sprechen, dann beziehen wir uns auf ein weites Begriffsverständnis von Learning Analytics im Sinne von Long et al. 2011: »Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.« (Long et al. 2011) Dabei gehen wir von einem konstruktivistischen Verständnis von Lernen aus, das auch mit LA-Ansätzen kompatibel ist (vgl. Leschke/Salden 2024: 198).

Realisiert wurde die KI-gestützte LA durch die Implementierung eines MCP-Servers, der mithilfe des Model Context Protocols (siehe Infokasten MCP) einer KI (Erprobung mit LLM Claude Opus 4.1) Werkzeuge in Form von Apps oder Datenbanken (in unserem Fall *AnnoPy*) zur Lösung komplexer Probleme zur Verfügung stellt. Wie genau dies umgesetzt wurde und welche Ergebnisse die Integration der KI erbrachte, wird im Folgenden dargelegt.

Infokasten MCP

Aktuelle KI-Modelle wie gpt-5 oder Claude Opus 4.5 sind leistungsstarke Werkzeuge zur Lösung von Problemen. Allerdings können sie nicht eigenständig mit anderen Anwendungen interagieren oder auf Echtzeitdaten zugreifen. KI-Modelle sind gewissermaßen Gehirne ohne Körper.[1] Eine Lösung für dieses Problem stellt MCP dar. Das Kürzel MCP steht für Model Context Protocol und definiert einen von Anthropic im November 2024 veröffentlichten Schnittstellenstandard (Client-Server-Architektur) für die Kommunikation zwischen KI-Modellen und externen Tools wie einer App oder einer Datenbank.[2]

Software-Entwickler können somit spezifizieren, auf welche Daten KI-Systeme zugreifen und wie sie verarbeitet werden sollen. MCP wird daher oft als »USB-C für KI-Anwendungen« bezeichnet. Die Verbindung zwischen Gehirn und Körper wird

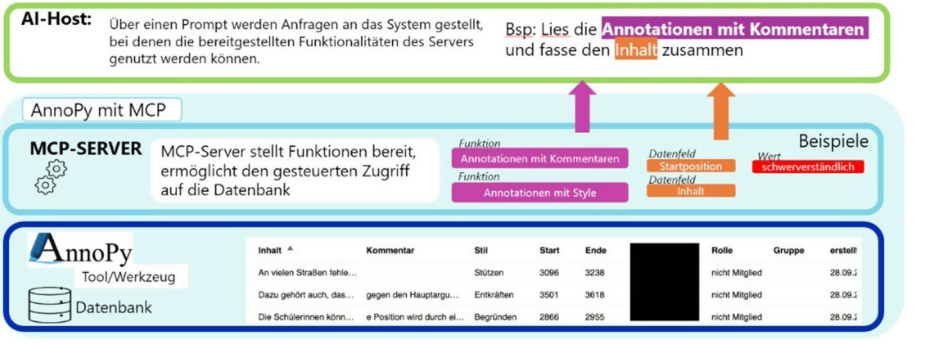
standardisiert und Interaktionen programmierbar, KI-Modelle können eigenständig agieren.

Quellen:
[1] Vgl. https://www.heise.de/blog/Maechtigere-KI-Systeme-mit-dem-Model-Context-Protocol-MCP-10363599.html vom 30.04.2025.
[2] Vgl. https://www.anthropic.com/news/model-context-protocol vom 25.11.2024.

Implementierung der KI-gestützten Auswertung

Aus technischer Sicht wurde eine textgenerierende KI (Claude) mit Hilfe der MCP-Technologie in *AnnoPy* integriert. Abb. 5 zeigt die Einbindung der relevanten technischen Bausteine der KI-basierten Learning Analytics in *AnnoPy*. Ein wesentlicher Baustein stellt der AI-Host dar. Hier werden über einen Prompt Anfragen an *AnnoPy* gestellt. In *AnnoPy* ist wiederum der MCP-Server integriert.

Abbildung 5: KI-gestützte Auswertung in AnnoPy.



Ein Prompt, z.B. in der Form »Lies die Annotationen mit Kommentaren und fasse den Inhalt zusammen«, der die Aufgabe der Hilfskraft ersetzten soll, wird über den AI-Host an den MCP-Server übertragen. Dieser stellt die für die Bearbeitung des Auftrags notwendigen Funktionen bereit, z.B. »Annotationen mit Kommentaren« oder »alle Annotationen mit Style«. Diese Funktionen sind serverseitig als Methoden implementiert und gestalten den Zugriff auf die entsprechenden Datenfelder der Datenbank, z.B. »Inhalt«, Kommentar« oder »Startposition«.

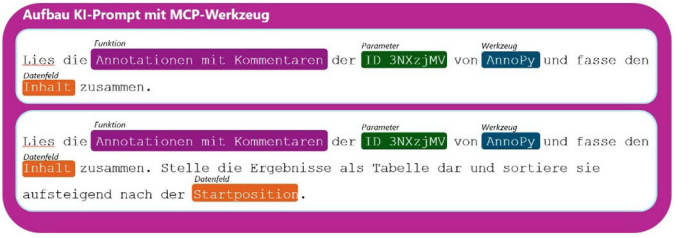
Das Besondere an der MCP-Technologie wird hieran deutlich. Die KI-Fähigkeiten, z.B. Inhalte von Texten zusammenzufassen, diese in Form von Tabellen darzustellen etc., können algorithmisiert eingebunden werden und ermöglichen

eine zielgerichtete Automatisierung von Prozessen (vgl. Infokasten MCP), die zuvor nur über komplexe und nicht automatisierbare ›Copy & Paste‹-Praktiken mit einem Chatbot erfolgen konnten.

Erste Ergebnisse des MCP-Einsatzes in *AnnoPy*

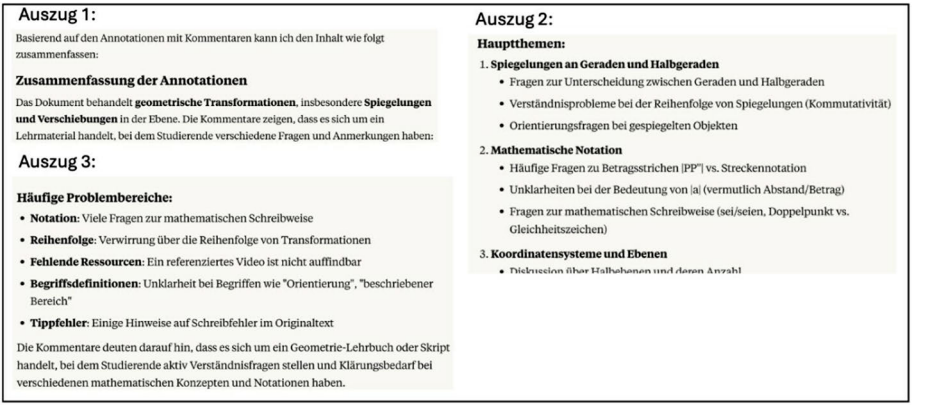
Im nächsten Schritt möchten wir die Auswertungsergebnisse vorstellen, die mittels der MCP-Technologie erzielt wurden. Ausgangspunkt sind zwei Prompts (vgl. Abb. 7) für die Auswertung der Kommentare, die die zuvor vorgestellte Anfrage konkretisieren:

Abb. 6: Beispielprompts mit MCP-Integration



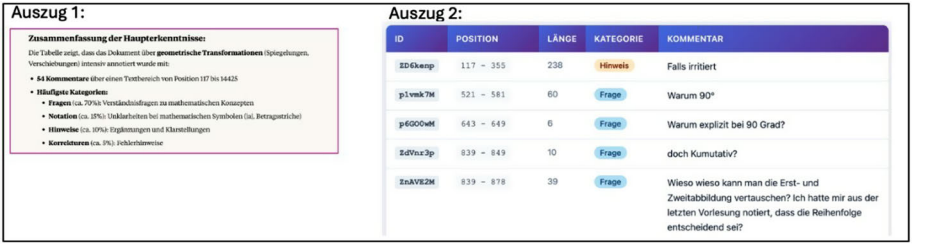
Der erste Prompt aus Abb. 6 bewirkt, dass die KI den Inhalt der Kommentare zunächst global zusammenfasst und im Anschluss daran Themenfelder sowie Problembereiche in den Kommentaren identifiziert und als geordnete Liste zur Verfügung stellt. Abb. 8 zeigt Auszüge aus der generierten Zusammenfassung. Es wird zunächst global zusammengefasst, welches Thema im Skript behandelt wird: geometrische Transformationen (Spiegelungen, Verschiebungen (Auszug 1)), danach werden dann Hauptthemen angeführt und dazu stichpunktartig ausgeführt, welche Fragen von den Studierenden genannt wurden (Auszug 2). Angeführt werden auch häufige Problembereiche, die die Studierenden genannt haben (Auszug 3) wie beispielsweise Probleme mit mathematischen Schreibweisen oder Begriffsdefinitionen. Diese Zusammenstellung bietet für Lehrende bereits gute Anknüpfungspunkte, um adaptive Lernangebote in der Vorlesung zu machen.

Abbildung. 7: Auszüge aus der generierten Zusammenfassung des 1. Prompts



Der zweite Prompt erweitert den ersten um ein definiertes Ausgabenformat (hier: die Tabellenform). Das Output beginnt ebenfalls mit einer globalen Zusammenfassung, wobei hier eine Gewichtung der Kategorien der Kommentare erfolgt und u.a. auch Auffälligkeiten hervorgehoben werden. Beispielsweise stellen die Studierenden sehr viele Fragen, geben aber wenige Hinweise in ihren Kommentaren. In der Liste (Auszug 2, Abb. 8) werden die einzelnen Kommentare der Studierenden in der rechten Spalte aufgeführt. Diesen Kommentaren werden entsprechende Textstellen aus dem Vorlesungsskript zugeordnet (Spalte 2: Position). Die KI nimmt eine Kategorisierung der Kommentare vor – wie Frage oder Hinweis (vgl. Spalte 4). Diese Kategorisierung wurde von der KI ohne explizite Aufforderung umgesetzt (vgl. Abb. 7)

Abbildung 8: Auszüge aus der generierten Zusammenfassung des 2. Prompts



Die Tabelle als Output ist auch insofern interessant, weil sie nicht als das finale Endprodukt betrachtet werden sollte, sondern als ein Zwischenprodukt, das algorithmisch ohne KI weiterverarbeitet werden kann. Die Verknüpfung der Kommentare mit den Textverweisen und den Stilen ermöglicht es der Lehrkraft, strukturiert

auf die Daten der Studierenden zuzugreifen. Ausgehend von einer solchen Tabelle können ausgewählte Kommentare der Studierenden zu entsprechenden Passagen des Skripts in der Lehrveranstaltung geordnet nach den jeweiligen Kategorien aufgegriffen und besprochen werden.

Fazit

Durch die Integration der KI-gestützten Auswertung in AnnoPy ist es im Vergleich zu einer rein algorithmischen und händischen Auswertung der Daten vergleichsweise schnell möglich, potenzielle Problembereiche, die Studierende beim Lesen von wissenschaftlichen Texten haben, aufzudecken und somit Lernsituationen auf der Basis dieser Datenauswertung besser an die Bedürfnisse der Lerngruppe in Bezug auf den Lerngegenstand anzupassen. Die KI-basierte Learning Analytics bietet daher Potenziale, die die Micro-Ebene (Lernende) auch als auch die Macro-Ebene (Lehrende) von Learning Analytics betreffen (vgl. Ifenthaler/Widanapathirana 2014).

Die Analysen zu den annotierten und kommentierten Textstellen können sozusagen ›auf Knopfdruck‹ durch die integrierte KI in Form von Tabellen mit entsprechenden Kategorisierungen oder Zusammenfassungen aufgezeigt werden. Es ist auch möglich, sich Ergebnisse grafisch visualisieren zu lassen, wenn entsprechende Prompts verwendet werden. Die Variabilität des Outputs ist unseres Erachtens von Vorteil, weil die Datendarstellung flexibel auf das jeweilige didaktische Szenario anpassbar ist.

Anzumerken ist, dass die LA-Daten nicht individuelle Daten einzelner Lernender abbilden und dies auch nicht intendiert ist. Es handelt sich um aggregierte Daten der spezifischen Lerngruppe bezogen auf einen konkreten Lerngegenstand und daher um eine spezifische Form von Learning Analytics. Diese Daten werden zu Reflexionszwecken, d.h. für die Anschlusskommunikation, in der gemeinsamen Lernsituation genutzt. Das Tool selbst bietet konzeptionell allen Beteiligten/Akteuren einen geregelten Zugang zu den Daten, d.h. sowohl die Lehrperson kann mit der gesamten Lerngruppe auf die aggregierten Daten zurückgreifen als auch einzelne Lerngruppen (ohne die Lehrkraft). Im Fokus steht dabei stets die soziale Interaktion und/oder die Kooperation der Peers (vgl. Leschke/Salden 2024: 190). Neben der gemeinsamen Reflexion bieten die Daten auch die Möglichkeit, bestimmte Bedarfe der Lerngruppe zu ermitteln und die Präsenzlehre entsprechend adaptiv zu gestalten – wie dies im bereits vorgestellten Vorlesungs-Beispiel erläutert wurde (vgl. Leschke/Salden 2024: 198).

Es muss selbstverständlich immer geprüft werden, wie die Güte der KI-generierten Ergebnisse im Vergleich zu einer händischen Auswertung ist. Im geschilderten Szenario war die Qualität durchaus zufriedenstellend.

Literatur

Alves, Rui A./Limpo, Teresa/Joshi, R. Malatesha (Hg.) (2020): Reading-Writing Connections. Towards Integrative Literacy Science, Cham: Springer.

Feilke, Helmuth (2001): Was ist und wie entsteht Literalität?, in: Pädagogik 6, S. 34–38.

Ifenthaler, Dirk/Widanapathirana, Chathuranga (2014): Development and Validation of a Learning Analytics Framework: Two Case Studies Using Support Vector Machines, in: Technology, knowledge and learning, Vol. 19, S. 221–240.

Kruse, Otto (2010): Lesen und Schreiben. Der richtige Umgang mit Texten im Studium, Konstanz: UVK Verlagsgesellschaft.

Leschke, Jonas/Salden, Peter (2024): »Didaktische Perspektiven auf Learning Analytics in der Hochschulbildung«, in: Peter Salden/Joas Leschke (Hg.): Learning Analytics und Künstliche Intelligenz in Studium und Lehre, Wiesbaden: Springer, S. 187–204.

Madlener-Charpentier, Karin/Behrens, Heike (2022): »Konstruktion(en) erst- und zweitsprachlichen Wissens: Lernprozesse und Steuerungsoptionen aus gebrauchsbasierter Perspektive«, in: Karin Madlener-Charpentier/Giulio Pagonis (Hg.): Aufmerksamkeitslenkung und Bewusstmachung in der Sprachvermittlung. Kognitive und didaktische Perspektiven auf Deutsch als Erst-, Zweit- und Fremdsprache Tübingen: Narr Francke Attempto, S. 31–66.

Portmann-Tselikas, Paul R./Schmölz-Eibinger, Sabine (2008): »Textkompetenz«, in: Fremdsprache Deutsch (39), S. 5–16.

Rezat, Sara/Rezat, Sebastian/Scholle, Oliver/Schulte, Carsten/Winkelnkemper, Felix (2024): »AnnoPy. Fachspezifische wissenschaftliche Textkompetenzen mit digitalen Medien in der Lehre fördern«, in: Herzig, Bardo et al. (Hg.): Lehrkräftebildung in der digitalen Welt. Zukunftsorientierte Forschungs- und Praxisperspektiven. Münster: Waxmann, S. 219–230.

Wie Learning Analytics und KI selbstreguliertes Lernen transformieren

Didaktische Potenziale videobasierter Lernumgebungen in der Lehrkräftebildung[1]

Manuel Oellers[2], Robin Junker[3], Jennifer Janeczko[4], Christina Dückers[5] und Manfred Holodynski[6]

Das didaktische Design von Blended Learning-Formaten zeichnet sich durch den Wechsel synchroner und asynchroner Lernphasen aus. Letztere ermöglichen zwar ein zeitflexibles und ortsunabhängiges Lernen, stellen jedoch zugleich erhöhte Anforderungen an die selbstregulativen Kompetenzen der Lernenden. Eine fehlende Unterstützung, wie unzureichendes Feedback, hindert dabei den Kompetenzerwerb. Der Beitrag erörtert, wie Unterstützungsbedarfe bereits im Lernprozess ermittelt werden und inwiefern Learning Analytics und KI die Lernenden in Selbstlernphasen durch formatives Feedback unterstützen können. Die Ausführungen schließen an unsere Studie an, in der Lernstrategien von Lehramtsstudierenden in einer digitalen, videobasierten Lernumgebung ausgewertet wurden. Darauf aufbauend werden didaktische Potenziale der Lernumgebung sowie konzeptionelle Ansätze der adaptiven Lernunterstützung aus der Praxis der Lehrkräftebildung diskutiert.

How learning analytics and AI are transforming self-regulated learning: Educational potential of video-based learning environments in teacher education

The instructional design of blended learning formats is characterized by alternating synchronous and asynchronous learning phases. While the latter enable learning that is flexible in terms of time and place, they also impose increased demands on learners' self-regulatory competencies. A lack of support, such as insufficient feedback, can hinder the development of these competencies. The article explores how learners' needs for support can be identified during the learning process and how learning analytics and AI can assist learners during

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID: 0000-0002-1580-8841
3 ORCID: 0000-0001-8558-2131
4 ORCID: 0000-0001-7253-1271
5 ORCID: 0000-0003-0706-3103
6 ORCID: 0000-0002-4791-4948

self-study phases by implementing formative feedback and assessment. The discussion builds on our study, which analyzes learning strategies of pre-service teachers in a digital, video-based learning environment. Building on these findings, the educational potential of such a learning environment, as well as conceptual approaches to adaptive learning support in teacher education, are discussed.

Herausforderungen und Perspektiven des zeitgemäßen selbstregulierten Lernens

Technologische Errungenschaften wie KI verändern nicht nur den Alltag, sondern auch die Art des Lernens auf eine nachhaltige Weise. Eine fortwährende Innovation von Lernumgebungen ist daher für zeitgemäßes Lernen unerlässlich, vor allem da *E-Learning* und auch der vielversprechende Einsatz von KI im Bildungsbereich an Bedeutung gewinnen (vgl. Caplanova et al. 2024; Crompton/Burke 2023, Tan et al. 2025).

Dabei vermögen klassische *Blended Learning*-Formate zwar die Vorteile von synchronen und asynchronen Lernphasen zu verbinden (vgl. De Bruijn-Smolders/Prinsen 2024; Cao 2023; Caplanova et al. 2024; Min/Yu 2023), doch eine unzureichende Unterstützung in einem isolierten, asynchronen Lernsetting kann zu erhöhten Abbruchquoten führen sowie die Lernmotivation und den Lernerfolg verringern (vgl. Rahmani et al. 2024; Prananto et al. 2025; Barz et al. 2024; Xie/Correia 2024).

Den Vorteilen des orts- und zeitunabhängigen Lernens stehen dabei zugleich hohe Anforderungen an die selbstregulativen Kompetenzen von Lernenden gegenüber (vgl. Zhang et al. 2024), die als ein entscheidender Faktor für das Erreichen oder Verfehlen der Lernziele gelten (vgl. Rahmani et al. 2024; Min/Yu 2023). Während in synchronen Lernphasen eine Lernunterstützung durch Dozierende erfolgen kann, erhalten Lernende in asynchronen Lernphasen hingegen häufig keine unmittelbaren Rückmeldungen zu ihren Lernstrategien oder Leistungen (vgl. Han et al. 2023; Hung et al. 2024). Auch Lehrenden bleibt oft ein Einblick in Lernprozesse verwehrt, so dass keine zielgerichtete Intervention erfolgen kann (vgl. Liu et al. 2025). Zwar bieten Lernmanagementsysteme (LMS) teils summative Indikatoren, jedoch bisweilen kaum adäquate Daten über die Lernverläufe und Hürden, die für den Lernerfolg relevant sind (vgl. Lobos et al. 2024; Faza/Lestari 2025). Wie können daher bereits im Lernprozess individuelle …

… Unterstützungsbedarfe und Leistungen ermittelt werden?
… Rückmeldungen und eine Lernunterstützung realisiert werden?

In Anlehnung an unsere Studie wird erörtert, inwiefern Unterstützungsbedarfe in Selbstlernphasen mit Learning Analytics (LA) in einer digitalen, videobasierten Lernumgebung erfasst und adressiert werden können (vgl. Oellers/Junker/Holodynski 2024). Darauf aufbauend wird ein konzeptioneller Ansatz dargelegt, der aufzeigt, inwiefern individuelle Rückmeldungen und KI-basiertes Feedback als adaptive Lernunterstützung begleitend in den Lernprozess implementiert werden können. Abschließend erfolgt eine Diskussion des didaktischen Potenzials aus der Praxis der Lehrkräftebildung im Hinblick auf eine Transformation des selbstregulierten Lernens (SRL).

Lernprozesse mit Learning Analytics begleiten

Ein bedeutsames Ziel der Lehrkräfteausbildung besteht in der Förderung professionellen Lehrkräftehandelns. Eine wesentliche Voraussetzung dafür stellt die professionelle Unterrichtswahrnehmung dar, als Kompetenz, lernrelevante Ereignisse im Unterricht erkennen, interpretieren und daraus begründete Handlungsentscheidungen ableiten zu können (vgl. Junker et al. 2020; Sherin/van Es 2009). Unterrichtsvideos gelten hierbei als wirksame Lerngelegenheiten zur Professionalisierung von Lehrkräften (vgl. Santagata et al. 2021; Sherin/van Es 2009). Durch eine theoriegestützte Analyse authentischer Unterrichtspraxis wird hierbei die Situierung erworbenen Wissens ermöglicht (vgl. Barth 2017; Santagata et al. 2007).

Zu diesem Zweck wurden in unserer Studie Unterrichtsvideos der Portale *ProVision* und *ViU: Early Science* der Universität Münster verwendet, die über das Meta-Videoportal *unterrichtsvideos.net* zugänglich sind (vgl. Oellers/Junker/Holodynski 2024). Die selektive Wahrnehmung lernrelevanter Unterrichtsereignisse stellt indes für Novizinnen und Novizen eine besondere Herausforderung dar, auch da Ereignisse unterschiedlich salient auftreten können (vgl. Telgmann/Müller 2023; Gabel et al. 2025). Die Einbindung in ein digitales Annotationswerkzeug als videobasierte Lernumgebung unterstützt die Wahrnehmung komplexer Unterrichtsereignisse sowie eine strukturierte schriftliche Analyse, begleitet durch formatives Feedback während der Analyse. Ein solches Feedback gilt als effektiver, wenn es zeitnah erfolgt (vgl. Shute 2008), sowie die Aufgabe, den Lernprozess und die Selbstregulation adressiert (vgl. Hattie/Timperley 2007). Auch die Qualität des Feedbacks ist entscheidend, wie bspw. eine Elaboration, Orientierung an fachlichen Kriterien und inhaltliche Spezifität (vgl. Prilop et al. 2021, Prilop et al. 2020). Die Bereitstellung formativen Feedbacks stellt insgesamt einen bedeutsamen Faktor für den Lernerfolg dar (vgl. Hattie/Timperley 2007; Hattie 2024; Henderson et al. 2019).

Damit solche Rückmeldungen im Lernverlauf adäquat erfolgen können, müssen zunächst individuelle Lernverläufe und -stände ermittelt werden (vgl. Morris et al. 2021; Zhang et al. 2023; Wiliam 2018). Learning Analytics (LA) stellt eine geeigne-

te Methode dar, um diese Lernprozesse in Echtzeit zu analysieren (vgl. Hoppe 2017; Lang et al. 2017; Knight/Buckingham Shum 2017; Siemens 2013). LA bezeichnet dabei das Erheben und Auswerten von (lernrelevanten) Daten aus Lernumgebungen mit dem Bestreben, Einblicke in Lernverläufe und Lernkontexte zu gewinnen und diese zu optimieren (vgl. Siemens 2013). Durch Verwendung gängiger *E-Learning*-Standards, wie *xAPI* oder *cmi5*, wird eine standardkonforme Erfassung verschiedener Lernerfahrungen ermöglicht, wie in diesem Beispiel exemplarisch für videobasierte Lernaktivitäten gezeigt wird (vgl. Mirriahi et al. 2016; Mirriahi et al. 2018; Oellers/Junker/Holodynski 2024).

Adaptive Lernunterstützung in videobasierten, digitalen Lernumgebungen

Lernrelevante Daten aus Lernumgebungen bilden das Fundament für eine adaptive Lernunterstützung (vgl. Ifenthaler/Yau 2020). Bereits in anderen Kontexten konnte zuvor gezeigt werden, dass eine LA-basierte Lernunterstützung und KI-basiertes Feedback einen positiven Effekt auf die selbstregulativen Kompetenzen und den Lernerfolg haben können (vgl. Tzimas/Demetriadis 2024; Luo et al. 2025; Seßler et al. 2025; Tepgec et al. 2025; Jacobsen/Weber 2025), wenngleich derzeit noch anzumerken ist, dass neben diesen Einzelbefunden bislang übergreifende systematische Metaanalysen fehlen (vgl. Leschke/Salden 2024).

Adaptive Lernunterstützungsmaßnahmen können dabei verschiedene Bereiche des Lernprozesses adressieren. Beim SRL planen und überwachen Studierende in verschiedenen Phasen eigenverantwortlich ihren Lernprozess (vgl. Zimmerman 2000; Pintrich 2000; Panadero 2017). Zum SRL zählen im Wesentlichen drei Phasen:

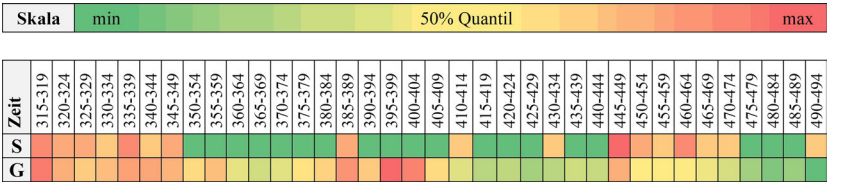
1. Vorbereitungsphase, in der Aufgaben und Ziele geplant werden
2. Handlungsphase, in der Lernstrategien angewendet werden
3. Reflexionsphase, in welcher der Lernprozess evaluiert wird

In diesen Phasen sind sowohl kognitive, metakognitive, motivationale, emotionale als auch behaviorale Aspekte des SRL von Relevanz (vgl. Ortube et al. 2024; Panadero 2017). Eine Lernunterstützung sollte diese Aspekte in asynchronen Lernsettings besonders fokussieren, da Lernende ihren Lernprozess ohne unmittelbare Rückmeldungen nicht effektiv überwachen oder ihr Lernverhalten anpassen können. Für videobasierte Lernaktivitäten, wie bspw. eine Unterrichtsanalyse mit einem Annotationswerkzeug, bieten sich dazu verschiedene Lernunterstützungsmaßnahmen an.

Prompting und Cueing in asynchronen Lernphasen

In digitalen, videobasierten Lernumgebungen kann der Betrachtungsverlauf einzelner Videosegmente mithilfe von LA erhoben werden (vgl. Oellers/Junker/Holodynski 2024), um auf diese Weise spezifische Unterstützungsbedarfe im Rahmen des SRL zu identifizieren und individuelles Feedback bereitzustellen. So ist beispielhaft für eine ausgewertete Lernaktivität festzustellen, dass der/die Studierende (S) gegenüber der gesamten Gruppe (G) kaum Aufmerksamkeit auf den Zeitraum von Minute 6:35 (Sekunde 395) bis Minute 6:45 (Sekunde 405) gelenkt und zudem größere Teile der Unterrichtssequenz eher oberflächlich betrachtet hat (vgl. Abb. 2).

Abbildung 1: Häufigkeit der Videosegmentwiederholung im Vergleich eines Studierenden (S) mit der Lerngruppe (G) für eine Unterrichtssequenz



Eine datenbasierte Lernunterstützung erlaubt es *Prompts* und *Cues* in den Lernprozess zu integrieren. *Prompting* bezeichnet hierbei einen gezielten Einsatz von handlungsauffordernden Impulsen (*Prompts*), um den Lernprozess zu unterstützen (vgl. Gentner et al. 2024; Gentner/Seufert 2020; Farrell et al. 2024), abgrenzend zum *KI-Prompting*, welches dazu dient, Instruktionen bspw. für *Large Lange Models* (LLM) zu formulieren. In videobasierten Lernaktivitäten kann dieses *Prompting* eine zentrale Rolle spielen, indem es Lernende dazu anregt, sich aktiv mit Unterrichtsereignissen oder der Analyse dieser auseinanderzusetzen. Dabei können kognitive und metakognitive Prompts eingesetzt werden. Exemplarische kognitive Prompts sind Aufforderungen zur Analyse vernachlässigter Sequenzen oder zur Beurteilung bereits vorgenommener Kodierungen. Metakognitive Prompts sind z. B. Aufforderungen zur Reflexion des eigenen Vorgehens oder zur Verwendung empfohlener Lernmaterialien. In mehreren Studien finden sich Hinweise, dass der kombinierte Einsatz den Lernerfolg und die Motivation der Lernenden erhöhen und kognitive Belastung reduzieren kann (vgl. Farrell et al. 2024; Zeitlhofer et al. 2023; Junker et al. 2025; Martin et al. 2022).

Cueing (vgl. Farrell et al. 2024) ist eine weitere Technik, bei der durch gezielte visuelle, auditive oder sprachliche Hinweise die Aufmerksamkeit der Lernenden gelenkt wird (vgl. van Gog 2014; Mayer/Moreno 2012), bspw. durch das Hervorheben

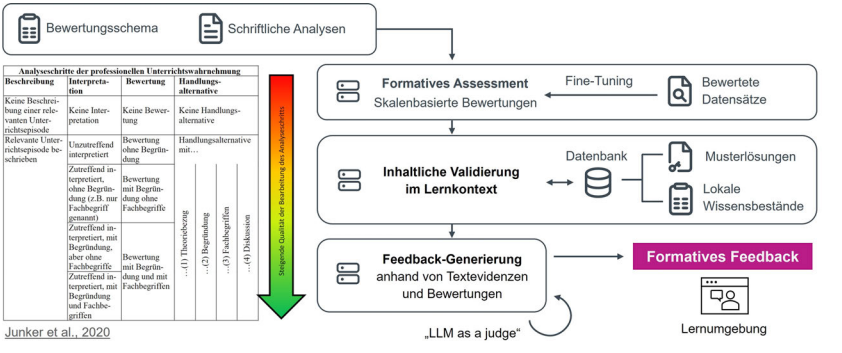
von lernrelevanten Ereignissen im Unterrichtsgeschehen, Sprungmarken oder Hinweise zu unberücksichtigten Analyseaspekten (vgl. Richter et al. 2016; Schneider et al. 2018). Zudem können zuvor erhobene Daten (wie Kodierungen im Annotationswerkzeug) in *Cues* wiederverwendet werden, um eine Übereinstimmung, Widersprüchlichkeit oder Kontroversität von inhaltlichen Kodierungen des Videomaterials zu signalisieren, falls bspw. mehrheitlich die gleichen, konträren oder vergleichbar anteilig gegensätzliche Kodierungen auftreten. Um auch bei schriftlichen, offenen Analysen unterstützen zu können, ist darüber hinaus ein formatives Feedback in diese Lernaktivitäten integrierbar.

KI-basiertes formatives Feedback in videobasierten Lernaktivitäten

Eine besondere Herausforderung für formatives Feedback im Lernprozess sind schriftliche, offene Aufgabenformate. In der Vergangenheit gab es bereits verschiedene Ansätze zur automatisierten Auswertung von Freitexten, wie etwa *Automated Essay Scoring* (AES), *Automated Writing Evaluation* (AWE) oder *Automated Written Feedback* (AWF) (vgl. Shi/Aryadoust 2024). Diese Ansätze basieren auf Verfahren zum maschinellen Lernen (ML) und/oder der Verarbeitung natürlicher Sprache (NLP).

Im Folgenden wird ein exemplarischer KI-Workflow vorgestellt, der sowohl Bewertungen und inhaltliche Validierungen vornimmt, sowie darüber hinaus ein elaboriertes Feedback zu schriftlichen Videoanalysen generiert. Zur Bewertung dieser Analysen wird ein entsprechend dafür entwickeltes Kodierschema (vgl. Junker et al. 2020) in ein maschinenlesbares Format übertragen und dient als kriterienorientierter Bewertungshorizont (*rubric-based assessment*), auch für zuvor bewertete Trainingsdatensätze.

Abbildung 2: KI-Workflow zur Generierung eines kriteriengeleiteten, formativen Feedbacks mit Textevidenzen als Explainable AI (XAI)-System



Während sich mit derzeitigem Stand für eine robustere, skalenbasierte Bewertung eher klassifizierende Transformer-Encoder-Modelle wie *DeBERTaV3*, *ModernBERT* oder *NeoBERT* eignen, ist für die Nutzung generativer KI zur Erstellung von elaboriertem Feedback zu berücksichtigen, dass klassische generative *Large Language Model* (LLM) zumeist auf einer eher allgemeinen Wissensbasis trainiert wurden. Ohne additive Maßnahmen sind diese zunächst nicht (ideal) geeignet, um domänenspezifische Inhalte kontextualisiert zu evaluieren und darauf aufbauend ein nützliches, fachliches inhaltliches Feedback für konkrete Aufgabenstellungen (z.B. mit kontextuellem Bezug zu einzelnen Videosequenzen) zu generieren.

Um die Konsistenz und Qualität des angestrebten Feedbacks für spezifische Aufgabenstellungen zu erhöhen, ist in diesem exemplarischen KI-Workflow ein *Fine-Tuning* des Modells vorgesehen, zumindest jedoch ein *Few-Shot-Learning*, entsprechend der Anzahl verfügbarer Trainingsdaten. Darüber hinaus findet ein Abruf kontextbezogener Informationen über das offene *Model Context Protocol* (MCP) statt. Dies ermöglicht den Zugriff auf Musterlösungen, zuvor bewertete Texte oder auch lokale Wissensbestände (u.a. Lernmaterial), um passendere Rückmeldungen generieren zu können. Relevante Inhalte werden dabei im Rahmen einer *Retrieval-Augmented-Generation* (RAG) über semantische Ähnlichkeiten ermittelt.

Als weitere Maßnahme werden Videosequenzen basierend auf Musterlösungen mit Wissensaussagen (*knowledge claims*) ausgezeichnet, die sodann bei den zu bewertenden Texten über *Natural Language Inference* (NLI) oder *Textual Entailment* (TE) evaluiert werden. Dabei wird den zu bewertenden Textpassagen zugeordnet, ob diese den Aussagen der Musterlösungen eher widersprechen (*contradicts knowledge claim*), die Aussagen eher implizieren (*entails knowledge claim*) oder dahingehend eher neutral sind. Ein anderweitiges LLM fungiert zur Prüfung von Zwischenschritten (»LLM as a judge«-Ansatz u.a. für *Guardrails*)

Auf dieser Basis wird formatives Feedback durch den Vergleich mit Musterlösungen generiert und zugleich anhand von Textevidenzen transparenter plausibilisiert. Der vorgestellte KI-basierte Workflow kombiniert dabei verschiedene Verfahren und erstellt schlussendlich eine strukturierte Rückmeldung mit konkreten Bezügen zu Textpassagen der zu bewertenden Zwischenlösung. Der KI-Workflow ist als *Microservice* konzipiert, mit dem Ziel, diesen in verschiedene Lernumgebungen zu integrieren, wie z.B. das Annotationswerkzeug und videobasierte Lernmodule, um formatives Feedback in selbstregulierten Lernprozessen unmittelbar bereitzustellen.

Fazit und Ausblick

Während herkömmliche, digitale Lernumgebungen häufig (noch) dazu dienen, Lerninhalte eher statisch bereitzustellen, ermöglichen LA- und KI-gestützte

Lernumgebungen neuartige didaktische Interventionen, insbesondere auch in asynchronen Lernsettings. Durch Auswertung lernrelevanter Daten kann das didaktische Potenzial unmittelbar im Lernprozess entfaltet werden. LA und KI ermöglichen auf diese Weise die Bereitstellung einer adaptiven Lernunterstützung in domänenspezifischen Lernkontexten, exemplarisch bspw. als *Prompts*, *Cues* und formatives Feedback, wie hier in der Praxis der Lehrkräftebildung für eine digitale, videobasierte Lernumgebung dargestellt. Über bestehende Befunde zur Vergleichbarkeit des Effekts von KI-basiertem Feedback gegenüber menschlichem Feedback (vgl. Kaliisa et al. 2025) hinaus, muss das Transferpotenzial für neue Anwendungskontexte künftig noch verstärkt beforscht werden. Grundlegend dafür sind eine sorgfältige Validierung und Evaluation der entwickelten Systeme, denn lernunterstützende Maßnahmen müssen gezielt und bedarfsorientiert wirken. Sie sollen den Lernprozess nicht überfrachten (z.B. in Form kognitiver Überlastung) und müssen zudem diversen Ansprüchen genügen, insbesondere inhaltlich, technisch und rechtlich (u.a. EU-DSGVO und EU-AI ACT). Es zeichnet sich jedoch bereits ab, dass LA- und KI-gestützte Lernumgebungen perspektivisch eine aussichtsreiche Chance auf messbare Erfolge bieten, selbstregulierte Lernprozesse lernförderlich zu transformieren.

Literatur

Barth, Victoria L. (2017): Professionelle Wahrnehmung von Störungen im Unterricht, Wiesbaden: Springer Fachmedien Wiesbaden.

Barz, Natalie/Benick, Manuela/Dörrenbächer-Ulrich, Laura/Perels, Franziska (2024): »Fostering self-regulated learning using synchronous or asynchronous digital learning environments: a latent profile analysis of pre-service teachers' individual differences«, in: Frontiers in Education 9, 1445182.

Cao, Wenwen (2023): »A meta-analysis of effects of blended learning on performance, attitude, achievement, and engagement across different countries«, in: Frontiers in Psychology 14, 1212056.

Caplanova, Anetta/Dunajeva, Jekatyerina/Rodriguez, Paula (2024): Digital transformation in blended learning environments – EENEE analytical report, Luxembourg: Publications Office of the European Union. https://data.europa.eu/doi/10.2766/061693 vom 29.08.2024.

Crompton, Helen/Burke, Diane (2023): »Artificial intelligence in higher education: the state of the field«, in: International Journal of Educational Technology in Higher Education 20.

De Bruijn-Smolders, M./Prinsen, Fleur R. (2024): »Effective student engagement with blended learning: A systematic review«, in: Heliyon 10, e39439.

Farrell, Meg/Martin, Monika/Böheim, Ricardo/Renkl, Alexander/Rieß, Werner/Könings, Karen D./van Merriënboer, Jeroen J. G./Seidel, Tina (2024): »Signaling cues and focused prompts for professional vision support: The interplay of instructional design and situational interest in preservice teachers' video analysis«, in: Instructional science 52, S. 879–917.

Faza, Ahmad/Lestari, Ilyana A. (2025): »Self-Regulated Learning in the Digital Age: A Systematic Review of Strategies, Technologies, Benefits, and Challenges«, in: The International Review of Research in Open and Distributed Learning 26(2), S. 23–58.

Gabel, Sylvia/Keskin, Özun/Gegenfurtner, Andreas (2025): »Comparing the effects of a specific task instruction and prompts on pre-service teachers' noticing of classroom management situations«, in: Zeitschrift für Erziehungswissenschaft 28, S. 105–123.

Gentner, Nadja/Respondek, Lisa/Seufert, Tina (2024): »Effects of short- and long-term prompting in learning journals on strategy use, self-efficacy, and learning outcomes«, in: Instructional science 52, S. 919–950.

Gentner, Nadja/Seufert, Tina (2020): »The Double-Edged Interactions of Prompts and Self-efficacy«, in: Metacognition and Learning 15, S. 261–289.

Han, Jaeyun/DiGiacomo, Daniela K./Usher, Ellen L. (2023): »College students' self-regulation in asynchronous online courses during COVID-19«, in: Studies in Higher Education 48(4), S. 1440–1454.

Hattie, John (2024): Visible learning 2.0, Baltmannsweiler: Schneider Verlag Hohengehren GmbH.

Hattie, John/Timperley, Helen (2007): »The Power of Feedback«, in: Review of Educational Research 77, S. 81–112.

Henderson, Michael/Ajjawi, Rola/Boud, David/Molloy, Elizabeth (Hg.) (2019): The Impact of Feedback in Higher Education. Improving Assessment Outcomes for Learners, Cham: Springer International Publishing; Imprint Palgrave Macmillan.

Hoppe, Heinz U. (2017): »Computational Methods for the Analysis of Learning and Knowledge Building Communities«, in: Lang, Charles/Siemens, George/Wise, Alyssa/Gasevic, Dragon (Hg.), Handbook of Learning Analytics, Society for Learning Analytics Research (SoLAR), S. 23–33.

Hung, Chih-Tsung/Wu, Shou-En/Chen, Yi-Hsien/Soong, Chen-Yeu/Chiang, Chien-Ping/Wang, Wei-Ming (2024): »The evaluation of synchronous and asynchronous online learning: student experience, learning outcomes, and cognitive load«, in: BMC medical education 24, S. 326.

Ifenthaler, Dirk/Yau, Jane Y.-K. (2020): »Utilising learning analytics to support study success in higher education: a systematic review«, in: Educational Technology Research and Development 68, S. 1961–1990.

Jacobsen, Lucas J./Weber, Kira E. (2025): »The Promises and Pitfalls of Large Language Models as Feedback Providers: A Study of Prompt Engineering and the Quality of AI-Driven Feedback«, in: AI 6, S. 35.

Janeczko, Jennifer M./Lehmkuhl, Alena/Junker, Robin/Zucker, Verena/Meschede, Nicola/Holodynski, Manfred (2025): »Promoting preservice teachers' multiperspective professional vision in virtual learning environments – On the effects of modeling videos and prompts during video-based lesson analysis«, in: Educational Technology Research and Development.

Junker, R./Janeczko, J./Lehmkuhl, A./Zucker, V./Holodynski, M./Meschede, N. (2025): »Modeling and prompting professional vision in a virtual learning environment: effects on pre-service teachers' cognitive load and motivation«, in: Education and information technologies 30, S. 19985–20010.

Junker, Robin/Rauterberg, Til/Möller, Kornelia/Holodynski, Manfred (2020): »Videobasierte Lehrmodule zur Förderung der professionellen Wahrnehmung von heterogenitätssensiblem Unterricht«, in: Herausforderung Lehrer*innenbildung – Zeitschrift Zur Konzeption, Gestaltung und Diskussion, 3, S. 236–255.

Kaliisa, Rogers/Misiejuk, Kamila/López-Pernas, Sonsoles/Saqr, Mohamed (2025): »How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions«, in: Educational Psychology, S. 1–32.

Knight, Simon/Buckingham Shum, Simon (2017): »Theory and Learning Analytics«, in: Lang, Charles/Siemens, George/Wise, Alyssa F./Gašević, Dragon/Merceron, Agathe (Hg.), Handbook of Learning Analytics, Society for Learning Analytics Research (SoLAR), S. 17–22.

Lang, Charles/Siemens, George/Wise, Alyssa F./Gašević, Dragon/Merceron, Agathe (Hg.) (2017): Handbook of Learning Analytics, Society for Learning Analytics Research (SoLAR).

Liu, Yan/Wang, Wei/Xu, Enwei (2025): »The Effectiveness of Learning Analytics-Based Interventions in Enhancing Students' Learning Effect: A Meta-Analysis of Empirical Studies«, in: SAGE Open 15.

Lobos, Karla/Cobo-Rendón, Rubia/Bruna Jofré, Daniela/Santana, Javiera (2024): »New challenges for higher education: self-regulated learning in blended learning contexts«, in: Frontiers in Education 9.

Luo, Jihao/Zheng, Chenxu/Yin, Jiamin/Teo, Hock H. (2025): »Design and assessment of AI-based learning tools in higher education: a systematic review«, in: International Journal of Educational Technology in Higher Education 22.

Martin, Monika/Farrell, Meg/Seidel, Tina/Rieß, Werner/Könings, Karen D./van Merriënboer, Jeroen J. G./Renkl, Alexander (2022): »Focused self-explanation prompts and segmenting foster pre-service teachers' professional vision – but only during training!«, in: International Journal of Educational Technology in Higher Education 19.

Mayer, Richard E./Moreno, Roxana (2012): »Techniques That Reduce Extraneous Cognitive Load and Manage Intrinsic Cognitive Load during Multimedia Learning«, in: Jan L Plass/Roxana Moreno/Roland Brünken (Hg.), Cognitive Load Theory, Cambridge University Press, S. 131–152.

Min, Wenhe/Yu, Zhonggen (2023): »A Systematic Review of Critical Success Factors in Blended Learning«, in: Education Sciences 13(5), S. 469.

Mirriahi, Negin/Jovanovic, Jelena/Dawson, Shane/Gašević, Dragan/Pardo, Abelardo (2018): »Identifying engagement patterns with video annotation activities: A case study in professional development«, in: Australasian Journal of Educational Technology 34.

Mirriahi, Negin/Liaqat, Daniyal/Dawson, Shane/Gašević, Dragan (2016): »Uncovering student learning profiles with a video annotation tool: reflective learning with and without instructional norms«, in: Educational Technology Research and Development 64(6), S. 1083–1106.

Morris, Rebecca/Perry, Thomas/Wardle, Lindsey (2021): »Formative assessment and feedback for learning in higher education: A systematic review«, in: Review of Education 9(3).

Oellers, Manuel/Junker, Robin/Holodynski, Manfred (2024): »Individual learning paths mastering teachers' professional vision«, in: Frontiers in Education 9.

Ortube, Alanze F./Panadero, Ernesto/Dignath, Charlotte (2024): »Self-Regulated Learning Interventions for Pre-service Teachers: a Systematic Review«, in: Educational Psychology Review 36(113).

Panadero, Ernesto (2017): »A Review of Self-regulated Learning: Six Models and Four Directions for Research«, in: Frontiers in psychology 8(422).

Pintrich, Paul R. (2000): »The Role of Goal Orientation in Self-Regulated Learning«, in: Boekaerts, Monique/Pintrich Paul R./Zeidner Moshe (Eds.), Handbook of Self-Regulation, Elsevier, S. 451–502.

Prananto, Kartika/Cahyadi, Surya/Lubis, Fitriani Y./Hinduan, Zahrotur R. (2025): »Perceived teacher support and student engagement among higher education students – a systematic literature review«, in: BMC psychology 13(112).

Prilop, Christopher N./Weber, Kira E./Kleinknecht, Marc (2020): »Effects of digital video-based feedback environments on pre-service teachers' feedback competence«, in: Computers in Human Behavior 102(110), S. 120–131.

Prilop, Christopher N./Weber, Kira E./Kleinknecht, Marc (2021): »The role of expert feedback in the development of pre-service teachers' professional vision of classroom management in an online blended learning environment«, in: Teaching and Teacher Education 99(4), 103276.

Rahmani, Amir M./Groot, Wim/Rahmani, Hamed (2024): »Dropout in online higher education: a systematic literature review«, in: International Journal of Educational Technology in Higher Education 21(9).

Richter, Juliane/Scheiter, Katharina/Eitel, Alexander (2016): »Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis«, in: Educational Research Review 17, S. 19–36.

Salden, Peter/Leschke, Jonas (2024): Learning Analytics und Künstliche Intelligenz in Studium und Lehre, Wiesbaden: Springer Fachmedien Wiesbaden.

Santagata, Rossella/König, Johannes/Scheiner, Thorsten/Nguyen, Ha/Adleff, Ann-Kristin/Yang, Xinrong/Kaiser, Gabriele (2021): »Mathematics teacher learning to notice: a systematic review of studies of video-based programs«, in: ZDM 53, S. 119–134.

Santagata, Rossella/Zannoni, Claudia/Stigler, James W. (2007): »The role of lesson analysis in pre-service teacher education: an empirical investigation of teacher learning from a virtual video-based field experience«, in: Journal of Mathematics Teacher Education 10, S. 123–140.

Schneider, Sascha/Beege, Maik/Nebel, Steve/Rey, Günter D. (2018): »A meta-analysis of how signaling affects learning with media«, in: Educational Research Review 23, S. 1–24.

Seßler, Kathrin/Bewersdorff, Arne/Nerdel, Claudia/Kasneci, Enkelejda (2025): Towards Adaptive Feedback with AI: Comparing the Feedback Quality of LLMs and Teachers on Experimentation Protocols. https://doi.org/10.48550/arXiv.2502.12842 vom 18.02.2025.

Sherin, Miriam/van Es, Elizabeth (2009): »Effects of Video Club Participation on Teachers' Professional Vision«, in: Journal of Teacher Education 60, S. 20–37.

Shi, Huawei/Aryadoust, Vahid (2024): »A systematic review of AI-based automated written feedback research«, in: ReCALL 36(2), S. 187–209.

Shute, Valerie J. (2008): »Focus on Formative Feedback«, in: Review of Educational Research 78, S. 153–189.

Siemens, George (2013): »Learning Analytics. The Emerge of a Discipline«, in: American Behavioral Scientist 57(10), S. 1380–1400.

Tan, Xiao/Cheng, Gary/Ling, Man H. (2025): »Artificial intelligence in teaching and teacher professional development: A systematic review«, in: Computers and Education: Artificial Intelligence 8, S. 100355.

Telgmann, Leonie/Müller, Katharina (2023): »Training & prompting pre-service teachers' noticing in a standardized classroom simulation – a mobile eye-tracking study«, in: Frontiers in Education 8.

Tepgec, Mustafa/Heil, Joana/Ifenthaler, Dirk (2025): »Feedback literacy matters: unlocking the potential of learning analytics-based feedback«, in: Assessment & Evaluation in Higher Education 50, S. 50–66.

Tzimas, Dimitrios E./Demetriadis, Stavros N. (2024): »Impact of Learning Analytics Guidance on Student Self-Regulated Learning Skills, Performance, and Satisfaction: A Mixed Methods Study«, in: Education Sciences 14, S. 92.

van Gog, Tamara (2014): »The Signaling (or Cueing) Principle in Multimedia Learning«, in: Richard E. Mayer (Hg.), The Cambridge Handbook of Multimedia Learning, Cambridge University Press, S. 263–278.

Wiliam, Dylan (2018): »Feedback«, in: Anastasiya A Lipnevich/Jeffrey K. Smith (Hg.), The Cambridge Handbook of Instructional Feedback, Cambridge University Press, S. 3–28.

Xie, Jiarui/Correia, Ana-Paula (2024): »The effects of instructor participation in asynchronous online discussions on student performance: A systematic review«, in: British Journal of Educational Technology 55, S. 71–89.

Zeitlhofer, Ines/Hörmann, Sandra/Mann, Bettina/Hallinger, Katharina/Zumbach, Joerg (2023): »Effects of Cognitive and Metacognitive Prompts on Learning Performance in Digital Learning Environments«, in: Knowledge 3(2), S. 277–292.

Zhang, Chengming/Hofmann, Florian/Kamran, Farrukh/Gläser-Zikuda, Michaela (2024): »Self-Regulated Learning of Pre-Service Teachers in Digital Learning Environments«, in: Meltzoff, Katherine/Mistretta, Sharon (Hg.), Self-Regulated Learning – Insights and Innovations, IntechOpen.

Zhang, Ke/Yilmaz, Ramazan/Ustun, Ahmet B./Karaoğlan Yilmaz, Fatma G. (2023): »Learning analytics in formative assessment: A systematic literature review«, in: Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi 14, S. 359–381.

Zimmerman, Barry J. (2000): »Self-Efficacy: An Essential Motive to Learn«, in: Contemporary educational psychology 25, S. 82–91.

Wie und von wem werden digitale Unterstützungsangebote zur Testung und Auffrischung schulischer Kompetenzen in der Hochschule genutzt?

Eine Evaluation anhand von Klickdaten und Befragungen[1]

Carolin Horsthemke[2], Elena Schmitt, Jakob Schwerter[3] und Hanna Gaspard[4]

*Kompetenzniveaus von Studienanfänger*innen variieren stark, weshalb Hochschulen den Ausgleich fehlender Kompetenzen durch ergänzende digitale Angebote zur Studienanforderungsbewältigung anstreben. Anhand von Klick-, Test- und Befragungsdaten im Lernmanagementsystem Moodle wurde über zwei Semester in mathematikintensiven Studiengängen untersucht, wie, von wem und mit welcher Motivation Online-Mathematiktests und -Kursangebote genutzt werden. Insgesamt zeigt sich eine intensive Testnutzung direkt nach Bereitstellung mit Präsenzbegleitung, jedoch geringe Kursnutzung im Semesterverlauf. Vergleiche zwischen Test-, Kursangebot- und Nicht-Nutzenden deuten auf eine stärkere Angebotsnutzung bei jüngeren Studierenden in niedrigeren Fachsemestern und mit geringeren Abbruchintentionen hin. Zudem nutzten Studierende mit häufigerer Nutzung des Fachveranstaltungs-Moodle-Kurses, höherem intrinsischen Wert und höherer Wichtigkeit von Mathematik sowie höheren testbezogenen Kosten das Angebot stärker.*

How and by whom are digital support offers for testing and refreshing school skills used in higher education? – An evaluation based on click data and surveys

Skill levels of university entrants vary greatly, which is why universities strive to compensate for missing skills with supplementary digital resources to meet study requirements. Using click, test, and survey data from the Moodle learning management system, it was examined how, by whom, and with what motivation online mathematics tests and courses are used in math-intensive study programs over two semesters. Overall, intensive test use is evident immediately after provision with on-site support, but course use is low during the semester.

1 Basiert auf einem Impulsbeitrag im Rahmen der Tagung.

2 ORCID:0009-0007-0919-666X

3 ORCID:0000-0001-5818-2431

4 ORICD:0000-0001-8830-8031

Comparisons between test users, course users, and non-users indicate greater use of the learning offer among younger students in lower semesters and with lower dropout intentions. In addition, students who used lecture-specific Moodle courses more frequently, who reported higher intrinsic and attainment value for mathematics as well as higher test-related costs made greater use of the offer.

Einleitung und theoretischer Hintergrund

Studienanfänger*innen variieren stark in ihren Kompetenzniveaus und Voraussetzungen für die Bewältigung von Studienanforderungen (Heublein et al. 2017: 47). Häufig sind fehlende Vorkenntnisse und dadurch entstehende Leistungsprobleme ein Grund für schlechtere Leistungen oder den Studienabbruch (Heublein et al. 2017: 13; Theune 2021: 34–36). In MINT-Studiengängen betrifft dies insbesondere mathematische Kompetenzen (Heublein et al. 2017: 56). Die Unterstützung bei Anforderungsbewältigung und Kompetenzausgleich ist daher Ziel der Hochschulbildung und des hier evaluierten Projekts »begleitendes Virtuelles nulltes Semester« (beVinuS).

Ergänzende digitale Lernangebote sollen Studierende unterstützen, da positive Zusammenhänge der Nutzung entsprechender Angebote mit den Klausurleistungen von Studierenden bekannt sind (Janson/Janke 2024: 11; Schwerter/Brahm 2024: 12). Eine Problematik in dieser Hinsicht ist allerdings, dass insbesondere Studierende mit schwächeren akademischen Leistungen und geringerer Motivation im Fach solche ergänzenden Lernangebote seltener in Anspruch nehmen (Chevalier/Dalton/Lührmann 2018: 338; Schwerter et al. 2025: 8).

Zur Untersuchung der Motivation für die Nutzung solcher Angebote kann die Situative Erwartungs-Wert Theorie als theoretischer Rahmen herangezogen werden. Deren zentrale Annahme ist, dass die Erfolgserwartung und Wertüberzeugungen von Lernenden das lern- und leistungsbezogene Verhalten beeinflussen (Eccles/Wigfield 2020: 2–3).

Fragestellung und Methode der Evaluation

In dieser Studie soll den Fragen nachgegangen werden, (1) wie ein ergänzendes digitales Lernangebot zur Aufarbeitung und Auffrischung schulischer Kompetenzen von Studierenden genutzt wird, (2) von wem es genutzt wird und (3) mit welcher Motivation.

Im Rahmen des beVinuS-Projekts wurde ein außercurriculares studienbegleitendes Angebot für MINT-Studiengänge erprobt und evaluiert (Abbildung 1). Das Lernangebot stand im Lernmanagementsystem der Hochschule in Form eines sepa-

raten beVinuS-Moodle-Kurses im Sommersemester 2024 (SoSe24) zunächst für die Zweitsemesterveranstaltungen der Elektrotechnik und Informationstechnik (ETIT) und Nachhaltige Energiesysteme (NES) zur Verfügung. Im Wintersemester 2024/25 (WiSe24/25) folgte die Erweiterung für das erste und dritte Semester sowie der Einsatz in Physik und Medizinphysik. Die Inhalte des Lernangebots umfassten Online-Mathematiktests und -Kursangebote zu schulmathematischen Voraussetzungen aller Grundlagenveranstaltungen der ersten drei Semester.

Das Lernangebot wurde zu Semesterbeginn in den Vorlesungen und Pflichtübungen der Studienfächer vorgestellt, mit der Gelegenheit, während der Veranstaltung am nun dauerhaft verfügbaren Online-Mathematiktest teilzunehmen und die kurzzeitig verfügbaren Befragungen auszufüllen. Kursangebote konnten flexibel im Anschluss an die Veranstaltung bearbeitet werden.

Zur Evaluation des Angebots wurden neben den Mathematiktestergebnissen Klick- und Befragungsdaten verwendet (Abbildung 1). Diese sollten Einblicke in die Angebotsnutzung, die Motivation und das Lernverhalten der Studierenden gewähren. Dazu wurde in den beVinuS-Moodle-Kursen am Anfang des SoSe24 und des WiSe24/25 ein Fragebogen für jeden der sechs Fachkurse sowie ein übergreifender Fragebogen platziert und die Klickdaten der Teilnehmenden über das Semester hinweg erfasst. Zur Vergleichbarkeit der Kursnutzung wurden zusätzlich die Klickdaten der Teilnehmenden in ihren Fachveranstaltungs-Moodle-Kursen erhoben.

Abbildung 1: Im Rahmen der Evaluation erfasste Daten in den verschiedenen Moodle-Kursen

Im SoSe24 wurde eine Stichprobe von $n = 135$ Teilnehmenden, im WiSe24/25 von $n = 478$ erzielt. Die Teilnehmenden waren im Durchschnitt 20.8 (SoSe24) bzw. 21.4 (WiSe24/25) Jahre alt und studierten im dritten Fachsemester (SoSe24: $M = 3.7$, WiSe24/25: $M = 3.4$), wobei die meisten im ersten bis dritten Semester waren. Nur wenige Studierenden waren weiblich (SoSe24: 15.6 %, WiSe24/25: 24.5 %) und Deutsch war zumeist Hauptsprache im Elternhaus (SoSe24: 75.0 %, WiSe24/25: 67.6 %).

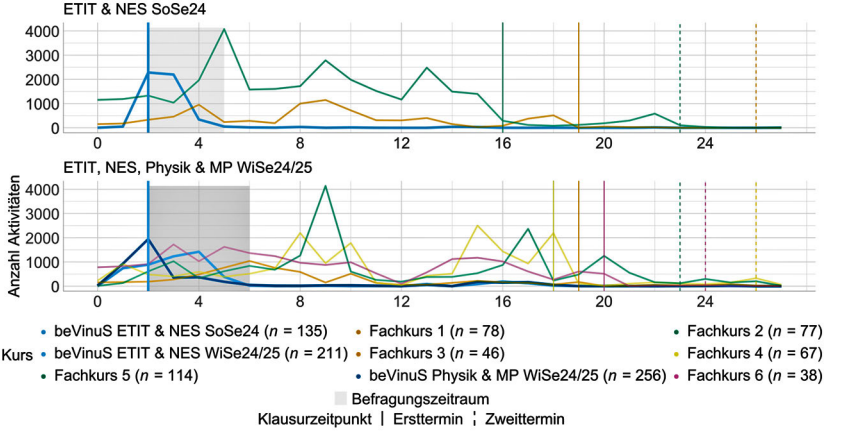
Zur Erfassung der Abbruchintentionen (Dresel/Grassinger 2013: 164) der Studierenden sowie ihrer Erwartungen (operationalisiert über das Selbstkonzept) und Wertüberzeugungen (unterteilt in intrinsischen Wert, Wichtigkeit, Nützlichkeit und Kosten) in Bezug auf Mathematik (Schwerter et al. 2022: 6) und den Test (Lauer-

mann et al. 2023: 108–109) wurden bereits validierte Skalen verwendet und teilweise adaptiert. Deren Antwortmöglichkeiten umfassten eine 6-stufige Likert-Skala von 1 = *stimme überhaupt nicht zu* bis 6 = *stimme voll und ganz zu*. Die Reliabilitäten der Skalen lagen in einem akzeptablen Bereich zwischen α = .61 und α = .94.

In Varianzanalysen und Post-hoc-Tests mit Bonferroni-Holm Adjustierung wurden Nutzende von *Test & Kursangebot,* Nutzende nur vom *Test* sowie *Nicht*-Nutzende verglichen. Von *Nicht*-Nutzenden lagen Daten vor, wenn sie an der Befragung teilgenommen hatten.

Ergebnisse

Abbildung 2: Klickaktivitäten pro Semesterwoche in den beobachteten Moodle-Kursen



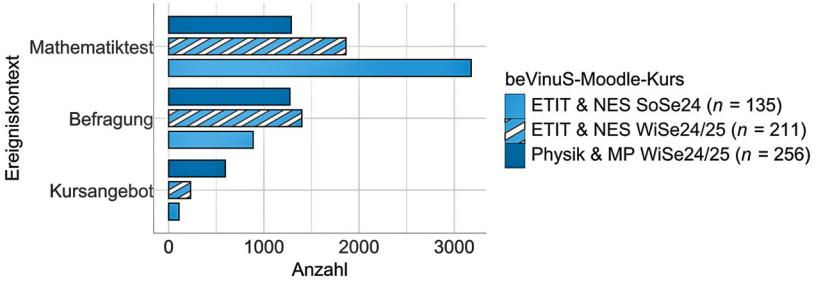
Anmerkung: ETIT = Elektrotechnik und Informationstechnik, NES = Nachhaltige Energiesysteme, SoSe24 = Sommersemester 2024, WiSe24/25 = Wintersemester 2024/2025, MP = Medizinphysik

Abbildung 2 zeigt die Klickaktivitäten der Teilnehmenden pro Semesterwoche in den beobachteten Moodle-Kursen. Die vertikalen Linien kennzeichnen darin jeweils die beVinuS-Testöffnungen beziehungsweise Klausurzeitpunkte der Fachkurse. Insgesamt ist für die beVinuS-Moodle-Kurse zu erkennen, dass die meisten Klickaktivitäten kurz nach der Testöffnung im Befragungszeitraum auftraten, während im Semesterverlauf kaum Klickaktivitäten zu verzeichnen sind. Im Gegensatz dazu gab es in den Fachveranstaltungs-Moodle-Kursen über das Semester

hinweg Klickaktivitäten, verstärkt zu Ereigniszeitpunkten wie Pflichtabgaben zur Klausurzulassung oder kurz vor den Klausuren.

Die am häufigsten angeklickten Inhalte des beVinuS-Online-Lernangebots (Abbildung 3) waren die beVinuS-Online-Mathematiktests, gefolgt von den Fragebögen. Die Kursangebote wurden hingegen deutlich seltener angeklickt.

Abbildung 3: Am häufigsten angeklickte Inhalte des beVinuS-Online-Lernangebots



Anmerkung: ETIT = Elektrotechnik und Informationstechnik, NES = Nachhaltige Energiesysteme, SoSe24 = Sommersemester 2024, WiSe24/25 = Wintersemester 2024/2025, MP = Medizinphysik

Beim Vergleich der Nutzenden-Gruppen (Tabelle 1) ergaben sich insbesondere für die *Nicht*-Nutzenden teils sehr kleine Gruppengrößen. Dennoch fanden sich im Gruppenvergleich für die *Nicht*-Nutzenden (marginal) signifikant höhere Fachsemester im SoSe24 ($p = .039$) sowie höhere Abbruchintentionen ($p = .051$) und Altersangaben ($p = .060$) im WiSe24/25. In der Mathematiktestleistung unterschieden sich die Gruppen nicht. Bezüglich der Fachveranstaltungs-Moodle-Kursnutzung zeigten sich für drei Fachkurse eine (marginal) signifikant häufigere Nutzung bei den Teilnehmenden, die das beVinuS-Angebot, insbesondere auch die Kursangebote, nutzten ($p = .004$ bis $p = .061$). Beim intrinsischen Wert ($p = .099$) und der Wichtigkeit ($p = .083$) von Mathematik fanden sich im WiSe24/25 höhere Ausprägungen bei *Test*-Nutzenden als *Nicht*-Nutzenden. Die *Test- & Kursangebot*-Nutzenden zeigten außerdem höhere testbezogene Kosten als die *Test*-Nutzenden ($p < .001$).

Tabelle 1: Vergleich der Nutzenden-Gruppen bezüglich der Nutzung des beVinuS-Online-Lernangebots

	SoSe24						WiSe24/25					
	Test & Kursangebot		Test		Nicht-Nutzende		Test & Kursangebot		Test		Nicht-Nutzende	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Mathematiktestleistung[1]	35	-0.1(0.8)	72	-0.2(1.0)	0	-	180	-0.1(1.0)	202	0.0(0.9)	0	-
Abbruchintentionen	9	2.4(1.0)	16	2.0(0.9)	2	3.5(0.5)	75	2.1(0.9)	76	1.9(0.8)	18	2.4(1.3)
Alter	10	19.9(2.3)	18	20.9(3.3)	3	22.7(0.6)	69	21.4(3.4)	68	20.9(3.5)	16	23.2(4.7)
Fachsemester	9	**2.4(0.9)**	18	**3.7(2.7)**	3	**7.0(4.4)**	68	3.7(4.3)	68	2.8(2.7)	16	4.1(3.5)
FK 1/3 Nutzungstage	29	**46.0(16.4)**	49	**37.2(21.3)**	0	-	14	27.1(25.7)	11	12.9(15.3)	4	17.0(17.8)
FK 2/4 Nutzungstage	31	47.9(21.5)	45	40.7(21.2)	1	4.0	26	**47.5(18.8)**	26	**47.7(26.9)**	4	**11.5(9.7)**
FK 5 Nutzungstage	-	-	-	-	-	-	53	**41.6(21.3)**	48	**30.4(18.1)**	5	**22.6(21.4)**
FK 6 Nutzungstage	-	-	-	-	-	-	16	52.4(29.1)	19	60.9(27.0)	3	28.0(30.8)
Mathematikmotivation												
Selbstkonzept	9	4.3(1.1)	15	4.4(0.9)	2	2.9(1.2)	75	4.2(1.2)	75	4.3(1.1)	18	3.9(1.1)
Intrinsischer Wert	9	4.1(1.1)	15	3.8(1.0)	2	4.4(0.9)	75	4.4(0.8)	75	4.4(1.0)	18	3.9(1.1)
Wichtigkeit	9	4.4(1.1)	15	3.7(1.1)	2	4.0(0.7)	75	4.2(1.1)	75	4.3(1.1)	18	3.7(1.4)
Nützlichkeit	9	4.5(1.0)	15	4.4(0.8)	2	4.8(1.8)	75	4.5(1.1)	75	4.5(1.1)	18	4.2(1.2)
Kosten	9	3.2(1.1)	15	3.1(1.1)	2	4.3(0.9)	-	-	-	-	-	-
Testmotivation												
Selbstkonzept	23	3.8(1.3)	23	3.8(1.6)	1	4.5(-)	79	3.5(1.0)	95	3.8(1.1)	3	4.2(0.3)
Intrinsischer Wert	23	4.3(1.1)	23	3.9(0.9)	1	3.5(-)	80	3.8(1.1)	95	3.5(1.2)	3	4.7(0.3)
Wichtigkeit/ Nützlichkeit	23	5.1(1.2)	23	5.1(1.0)	1	4.0(-)	79	4.6(0.9)	95	4.6(0.9)	3	5.2(0.3)
Kosten	23	2.6(1.1)	23	2.5(1.0)	1	1.0(-)	79	**3.5(1.1)**	95	**2.8(1.2)**	3	2.3(0.3)

Anmerkung: [1]z-Standardisierung der Werte; FK = Fachkurs; Fett gedruckte Werte kennzeichnen signifikant unterschiedliche Gruppen zum Niveau von $p < .05$;

Diskussion

Die insgesamt eher geringe Nutzung des außercurricularen digitalen Lernangebots von beVinuS ist vergleichbar mit Befunden zu alternativen Möglichkeiten der Kompetenzauffrischung schulischer Mathematikinhalte (Falk/Marschall 2021: 351; Tieben 2019: 1185). Außerdem konnten anhand der Klickaktivitäten in den Moodle-Kursen Verhaltensmuster in Abhängigkeit bestimmter Ereignisse beobachtet werden, wie ein erhöhtes Klickverhalten im beVinuS-Befragungszeitraum sowie zu Pflichtabgaben in den Fachkursen und Klausuren (siehe auch Schwerter/Brahm 2024: 9; von Keyserlingk et al. 2023: 141).

Darüber hinaus könnten sich höhere testbezogene Kosten bei *Test- & Kursangebot*-Nutzenden möglicherweise damit erklären lassen, dass die im Angebot enthaltenen Selbsttestoptionen Defizite aufzeigten und bewusst machten (Tieben 2019: 1181–1182), was zur weiteren Nutzung des Lernangebots angeregt haben könnte. Die Tendenz, bei höherer Fachveranstaltungs-Moodle-Kursnutzung im Vergleich zu Mitstudierenden neben den Tests auch Kursangebote zu nutzen, deutet auf ein generell höheres Engagement dieser Studierendengruppe hin. Höhere Ausprägungen des intrinsischen Werts und der Wichtigkeit bei *Test*-Nutzenden im Vergleich zu *Nicht*-Nutzenden entsprechen häufig gefundenen Mustern in Bezug auf eine höhere Nutzung von freiwilligen Lernangeboten durch motivierte Studierende (Schwerter et al. 2025: 8–10, 13).

Des Weiteren unterlagen die Lernangebotsnutzung und Begleitstudienteilnahme einer Selbstselektion durch vermutlich eher motiviertere Studierende. Dies zeigte auch die sehr kleine Gruppe *Nicht*-Nutzender im Rahmen der Studie. Aufgrund der zum Teil sehr kleinen Gruppengrößen sind die Ergebnisse vorsichtig zu interpretieren und mit einer größeren Stichprobe erneut zu untersuchen. Die Erfassung der Klickdaten war auf die Moodle-Kurse beschränkt, wodurch externes Lernverhalten nicht beobachtbar war. Die Klickdaten geben außerdem keine Auskunft darüber, ob die Inhalte nur angeklickt oder tatsächlich gelernt wurden.

Allerdings konnten reale Lernsituationen durch die Anknüpfung des Lernangebots an Fachveranstaltungen sowie die Kombination von Klick- und Befragungsdaten im Rahmen von Learning Analytics beobachtet werden (von Keyserlingk et al. 2023: 130). Somit konnten die Ergebnisse in das Projekt und die Fakultäten rückgekoppelt werden, was zu Anstößen für die Weiterentwicklung von Unterstützungsangeboten führte. Aus der Beobachtung, dass die Angebote insgesamt wenig und vor allem bei Präsenzbegleitung genutzt wurden, resultiert für das Projekt, dass Unterstützungsangebote zukünftig in Form von studienbegleitenden Präsenzformaten organisiert werden sollten. Die Studie dokumentierte anhand der kombinierten Analyse von Moodle-Nutzungsdaten und Studierendenbefragungen einen konkreten Implementierungsansatz für datenbasierte Evaluationsmethoden in der Hochschulbildung sowie deren Anwendbarkeit zur Bewertung digitaler Lernangebote.

Danksagung

Das Projekt beVinuS.nrw wurde durch das Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen (MKW) gefördert und an den Hochschulen BU Wuppertal, RWTH Aachen und TU Dortmund umgesetzt. Zudem wurde diese Arbeit durch das Projekt FAIR (PROFILNRW-2020-068) aus dem Programm »Profilbildung 2020« des MKWs unterstützt. Außerdem gilt unser Dank der studentischen Hilfskraft Lisa Larraß.

Literatur

Chevalier, Arnaud/Dalton, Peter/Lührmann, Melanie (2018): »›Making it count‹: Incentives, student effort and performance«, in: Journal of the Royal Statistical Society Series A: Statistics in Society 181(2), S. 323–349.

Dresel, Markus/Grassinger, Robert (2013): »Changes in achievement motivation among university freshmen«, in: Journal of Education and Training Studies 1(2), S. 159–173.

Eccles, Jacquelynne S./Wigfield, Allan (2020): »From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation«, in: Contemporary Educational Psychology 61, 101859, S. 1–13.

Falk, Susanne/Marschall, Maximiliane (2021): »Abbruch des Erststudiums bei MINT-Studierenden: Welche Rolle spielen Informations- und Unterstützungsangebote bei Studienbeginn?«, in: Martin Neugebauer/Hans-Dieter Daniel/Andrä Wolter (Hg.), Studienerfolg und Studienabbruch, Wiesbaden, Deutschland: Springer VS, S. 345–368.

Heublein, Ulrich/Ebert, Julia/Hutzsch, Christopher/Isleib, Sören/König, Richard/Richter, Johanna/Woisch, Andreas (2017): »Motive und Ursachen des Studienabbruchs an baden-württembergischen Hochschulen und beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher«, Hannover, Deutschland: DZHW Projektbericht. https://www.dzhw.eu/pdf/21/studienabbruchquoten_absolventen_2016.pdf

Janson, Marc P./Janke, Stefan (2024): »The influence of e-learning on exam performance and the role of achievement goals in shaping learning patterns«, in: International Journal of Educational Technology in Higher Education 21(1), 56.

Lauermann, Fani/Benden, Daria K./DeVries, Jeffrey M./Heitzer, Johanna (2023): »Students' use of (online) self-assessments in math: Interindividual differences in students' test engagement and test performance in math-intensive study programs«, in: Nele McElvany/Alyssa L. Grecu/Ramona Lorenz/Michael Becker/Charlotte Dignath/Hanna Gaspard/Fani Lauermann (Hg.), Jahrbuch der Schul-

entwicklung: 50 Jahre Schulentwicklung – Leitthemen der empirischen Bildungsforschung, Band 23, Weinheim, Basel: Beltz Juventa, S. 97–127, 961277.

Schwerter, Jakob/Brahm, Taiga (2024): »Voluntary e-learning exercises support students in mastering statistics«, in: Technology, Knowledge and Learning 29, S. 1437–1474.

Schwerter, Jakob/Dimpfl, Thomas/Bleher, Johannes/Murayama, Kou (2022). »Benefits of additional online practice opportunities in higher education«, in: Internet and Higher Education 53, 100834.

Schwerter, Jakob/Lauermann, Fani/Brahm, Taiga/Murayama, Kou (2025): »Differential use and effectiveness of practice testing: Who benefits and who engages?«, in: Learning and Individual Differences 123, S. 1–17, 102761.

Theune, Katja (2021): »Determinanten und Modelle zur Prognose von Studienabbrüchen«, in: Martin Neugebauer/Hans-Dieter Daniel/Andrä Wolter (Hg.), Studienerfolg und Studienabbruch, Wiesbaden, Deutschland: Springer VS, S. 19–40.

Tieben, Nicole (2019): »Brückenkursteilnahme und Studienabbruch in Ingenieurwissenschaftlichen Studiengängen«, in: Zeitschrift für Erziehungswissenschaft 22, S. 1175–1202.

von Keyserlingk, Luise/Lauermann, Fani/Yu, Renzhe/Rubach, Charlott/Arum, Richard (2023): »What can digital trace data tell us about post-secondary students' academic success? An overview of the literature and an illustrative example «, in: Nele McElvany/Alyssa L. Grecu/Ramona Lorenz/Michael Becker/Charlotte Dignath/Hanna Gaspard/Fani Lauermann (Hg.), Jahrbuch der Schulentwicklung: 50 Jahre Schulentwicklung – Leitthemen der empirischen Bildungsforschung, Band 23, Weinheim, Basel: Beltz Juventa, S. 128–149.