

EP

Empirische Pädagogik – 2024 – 38 (1)

Roland Happ, Kristina Kögler, Jacqueline
Schmidt & Marc Egloffstein (Hrsg.)
Lehren und Lernen mit und über Künstliche
Intelligenz in der Aus- und Weiterbildung

Empirische Pädagogik

38. Jahrgang 2024

1. Heft

Herausgeber

Zentrum für Empirische Pädagogische Forschung (zepf)
Bürgerstraße 23, 76829 Landau/Pfalz
Telefon: +49 6341 280 32165, Telefax: +49 6341 280 32166
E-Mail: sekretariat@zepf.uni-landau.de
Homepage: www.zepf.eu

Verlag

Empirische Pädagogik e. V.
Bürgerstraße 23, 76829 Landau/Pfalz
Telefon: +49 6341 280 32180, Telefax: +49 6341 280 32166
E-Mail: info@vep-landau.de
Homepage: www.vep-landau.de

Druck

Sowa Piaseczno

Zitiervorschlag

Happ, R., Kögler, K., Schmidt, J. & Egloffstein, M. (Hrsg.). (2024). Lehren und Lernen mit und über Künstliche Intelligenz in der Aus- und Weiterbildung (Empirische Pädagogik, 38 (1), Themenheft). Landau: Verlag Empirische Pädagogik.

Alle Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung, werden vorbehalten. Kein Teil des Werks darf in irgendeiner Form (durch Fotografie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlags reproduziert oder unter Verwendung elektronischer Systeme verbreitet werden.

ISSN 0931-5020

ISBN 978-3-944996-97-4

© Verlag Empirische Pädagogik, Landau 2024

<https://doi.org/10.62350/RZIP6202>

Inhalt

Editorial

Roland Happ, Kristina Kögler, Jacqueline Schmidt & Marc Egloffstein Lehren und Lernen mit und über Künstliche Intelligenz in der Aus- und Weiterbildung	6
---	---

Thementeil

Jacqueline Schmidt & Roland Happ Validierung eines Tests zu Grundlagenwissen über Künstliche Intelligenz von angehenden Lehrkräften	9
Jule Hangen & Eveline Wuttke Künstliche Intelligenz verstehen lernen – Ergebnisse eines Trainingsprogramms zur Förderung von Startup-Fähigkeiten im Bereich KI	26
Sabine Seufert, Lukas Spirgi, Jan Delcker, Joana Heil & Dirk Ifenthaler Umgang mit KI-Robotern: maschinelle Übersetzer, Textgeneratoren, Chatbots & Co – Eine empirische Studie bei Erstsemester-Studierenden	47
Josef Guggemos, Jacqueline Schmidt & Roland Happ Eine Frage von Macht: Einstellungen angehender Lehrkräfte zu den ethischen Grundsätzen des Einsatzes künstlicher Intelligenz in der Bildung	73
Marc Egloffstein, Kristina Kögler & Dirk Ifenthaler Evidenzgestützte Entwicklung von onlinebasierten Lernangeboten zu Künstlicher Intelligenz in der beruflichen Bildung: Stakeholder- Perspektiven und Implementierung	98
Julia Pargmann, Anna Leube, Florian Berding, Elisabeth Riebenbauer, Karin Rebmann, Andreas Slopinski & Michael Gillert Die elektronisch-didaktische Assistenz (EDDA) zur Unterrichtsplanung auf Basis künstlicher Intelligenz: Funktionsweise, Anwendungsbereiche und Forschungsperspektiven	118

Allgemeiner Teil

Joel Guttke & Raphaela Porsch	
Kognitive Interviews als Methode der Inhaltsvalidierung von	
Fragebogenitems zur Erfassung kognitiver Aktivierung im	
Engischunterricht der Grundschule.....	147

Rezension

Wolfgang Klug	
Melanie Misamer (2023): Machtsensibilität in der Sozialen Arbeit:	
Grundwissen für reflektiertes Handeln	168
Call for Guest Editors	173
Impressum	175

Contents

Articles

Jacqueline Schmidt & Roland Happ	
Validation of a test on prospective teachers' basic knowledge about artificial intelligence	9
Jule Hangen & Eveline Wuttke	
Learning about artificial intelligence – results from a training program to promote startup skills in the AI field.....	26
Sabine Seufert, Lukas Spirgi, Jan Delcker, Joana Heil & Dirk Ifenthaler	
Interaction with AI robots: machine translators, text generators, chatbots & co - an empirical study among first-year students.....	47
Josef Guggemos, Jacqueline Schmidt & Roland Happ	
A matter of power: prospective teachers' attitudes towards the ethical principles of artificial intelligence use in education	73
Marc Egloffstein, Kristina Kögler & Dirk Ifenthaler	
Evidence-based development of online learning resources on Artificial Intelligence in vocational education and training: Stakeholder perspectives and implementation.....	98
Julia Pargmann, Anna Leube, Florian Berding, Elisabeth Riebenbauer, Karin Rebmann, Andreas Slopinski & Michael Gillert	
Electronic didactic assistance (EDDA) for lesson planning based on artificial intelligence: functionality, areas of application and research perspectives.....	118

Topic-centered

Joel Guttke & Raphaela Porsch	
Assessing the content validity of questionnaire items on cognitive activation in English language teaching at primary level through cognitive interviewing	147

Editorial

Roland Happ, Kristina Kögler, Jacqueline Schmidt & Marc Egloffstein

Lehren und Lernen mit und über Künstliche Intelligenz in der Aus- und Weiterbildung

Konzepte und Anwendungen der Künstlichen Intelligenz (KI) haben in jüngerer Zeit im Bildungskontext und insbesondere auch in der beruflichen Aus- und Weiterbildung immens an Relevanz und Sichtbarkeit gewonnen. Dabei werden der Einfluss von KI-gestützten Systemen auf zahlreiche Berufsbilder einerseits, andererseits aber auch die Potenziale und Herausforderungen von KI-Anwendungen für die Gestaltung von Lehr- und Lernprozessen intensiv diskutiert. Es lassen sich somit in der Auseinandersetzung mit KI im beruflichen Bildungskontext mindestens zwei wichtige Perspektiven ausmachen, die beide umfassende Fragen auf verschiedenen Ebenen nach sich ziehen: Zum einen geht es um KI als zunehmend bedeutenden Bildungsinhalt, ‚über‘ den es zu lernen gilt, zum anderen um KI als Methodik, ‚mit‘ der sich Lehr-Lern- und Arbeitsprozesse verändert denken und gestalten lassen. Entsprechende Forschungs- und Entwicklungsvorhaben, die sich mit den KI-induzierten Veränderungen in Aus- und Weiterbildung befassen, sind angesichts der ausgesprochen dynamischen Entwicklungen der jüngeren Zeit zunehmend mit der Herausforderung konfrontiert, anschlussfähig an den jeweils aktuellen Erkenntnis- und Implementierungsstand zu sein und gleichzeitig mit Blick auf Nachhaltigkeitsfragen auch Erkenntnisse zu generieren, die zumindest mittelfristig Bestand haben. Dies bedeutet insbesondere für Vorhaben, die über generische Fragen und Zugänge hinausgehen, dass sie eher den Charakter einer Momentaufnahme haben. Vor diesem Problemhintergrund widmet sich die vorliegende Ausgabe der Zeitschrift Empirische Pädagogik aktuellen Forschungsansätzen zu KI im berufsbildenden Bereich. Mit den Beiträgen des Themenheftes werden drei thematische Akzente adressiert:

(1) Ansätze zur Erfassung und Förderung KI-bezogener Kompetenzfacetten
Schmidt und Happ berichten von der Validierung eines Testinstruments zur Erfassung des Grundlagenwissens zu KI von angehenden Lehrkräften im berufsbildenden Bereich. Dabei werden die Analysen zur Beurteilung der Validierungsaspekte "Testinhalt" und "Beziehung zu anderen Merkmalen" diskutiert. Auf Basis der Befunde können theoretisch postulierte Annahmen zum KI-Wissen von (angehenden) Lehrkräften empirisch untermauert werden, wodurch ein wichtiger Beitrag zur Nutzbarkeit des entwickelten Instruments geleistet wird.

Hangen und Wuttke nehmen in ihrer Mixed-Methods Evaluationsstudie Wissen, Motivation sowie Überzeugungen und Einstellungen von (angehenden) Gründerinnen und Gründern von Startups in den Blick. Im Prä-Post-Vergleich nimmt das selbstberichtete Wissen über Unternehmensgründungen zu. Die Evaluationsergebnisse zeigen allerdings keine signifikante Wissensveränderung über Anwendungen der KI. In Bezug auf Gründungsaspekte kann das Trainingsprogramm daher als effektiv angesehen werden, weniger jedoch im Hinblick auf KI.

(2) Ethische Herausforderungen des Einsatzes von KI in Bildungsprozessen.

Den Umgang mit KI-Robotern untersuchen Seufert, Spirgi, Delcker, Heil und Ifenthaler in einer empirischen Studie mit Studierenden im ersten Semester (N = 636). In einem weiten Verständnis fassen die Autor*innen darunter Systeme mit menschenähnlicher Leistung (z. B. Übersetzer, Schreibassistenten) sowie Systeme mit menschenähnlicher Erscheinung (z. B. Chatbots, Avatare). Ausgehend von sieben Anwendungsfällen werden die Nutzungshäufigkeit sowie die ethische Beurteilung in den Blick genommen. In der Gesamtschau überwiegen dabei die wahrgenommenen Risiken gegenüber den identifizierten Chancen. Die Nutzungshäufigkeit von KI-Robotern war zudem im Erhebungszeitraum – kurz vor der Veröffentlichung von ChatGPT – als eher gering einzustufen.

Guggemos, Schmidt und Happ untersuchen die Einstellungen angehender Lehrkräfte zu den ethischen Grundsätzen des Einsatzes von KI im Bildungswesen. In der Studie mit N = 90 angehenden Lehrkräften erweist sich das postulierte Messmodell zur Bewertung der Einstellung als reliabel und valide. Eine latente Profilanalyse führt zu drei von Kontextvariablen unabhängigen Profilen mit jeweils unterschiedlichen Einstellungsstrukturen, die sich vor allem in der Einschätzung unterscheiden, ob es der KI erlaubt sein sollte, harte Macht über Lehrkräfte und Lernende auszuüben.

(3) Evidenzgestützte Entwicklung von KI-bezogenen Lehr-Lern-Ressourcen

Egloffstein, Kögler und Ifenthaler beschreiben die evidenzgestützte Entwicklung von Online-Lernangeboten zu KI in der beruflichen Bildung. Im Zentrum steht dabei eine qualitative Studie zur Zielgruppen- und Kontextanalyse, in der N = 48 Akteur*innen aus der beruflichen Bildung zu KI-bezogenen Aspekten befragt wurden. Auf Basis der Ergebnisse wurden Impulse für die Gestaltung von AI_VET, einer Serie von vier Online-Kursen auf der Plattform des KI-Campus, abgeleitet. Erste Evaluationsergebnisse deuten auf eine differenzierte Nutzung der implementierten Kursbausteine sowie auf Akzeptanz auf Seiten der Lernenden hin.

Pargmann et al. stellen eine KI-Plattform zur Unterstützung der Planungskompetenz von Unterricht vor. Die elektronisch-didaktische Assistenz (EDDA) kann für die Analyse von Unterrichtsentwürfen und -materialien im Studium, im Vorbereitungsdienst und in der schulischen Berufstätigkeit verwendet werden. EDDA stellt Rückmeldungen zur Umsetzung ausgewählter didaktischer Merkmale der Unterrichtsplanung

bereit und liefert dabei Hinweise zur Reflexion und Weiterentwicklung der Entwürfe. Darüber hinaus kann EDDA als Ausgangsbasis für weitere Forschungen zum Lehren und Lernen mit KI dienen.

Anhand der drei Themenakzente lassen sich wichtige Spannungsfelder nachzeichnen, in denen sich die Diskussion um KI in der beruflichen Bildung aktuell bewegt. Das Spannungsfeld zwischen einer hohen Entwicklungsdynamik und dem Streben nach Nachhaltigkeit offenbart sich etwa besonders dann, wenn es darum geht, Wissen über KI in einem Testverfahren zu operationalisieren. Die ersten beiden Artikel verdeutlichen, wie schnell die Entwicklungen im Themenfeld KI das Design der Messinstrumente, insbesondere auf der Inhaltsebene, beeinflussen. Es bedarf der ständigen Anpassung von Testinstrumenten, um valide Testwertinterpretationen zuzulassen. Hierbei hat sich die interdisziplinäre Zusammenarbeit (bspw. mit der Wirtschaftsinformatik) als hilfreich und notwendig erwiesen. Dass die technischen Möglichkeiten immer auch in Bezug zu ethischen Herausforderungen abgewogen werden müssen, ist die Quintessenz der Beiträge aus Bereich zwei. Aus beiden Beiträgen wird deutlich, dass der stärkere Verbreitungsgrad des Chatbots ChatGPT auch auf non-kognitive Facetten eine Wirkung haben sollte. Für die Zukunft bieten sich mit den bestehenden Messinstrumenten zu den non-kognitiven Facetten auch Kohortenvergleiche an (bspw. Stichprobe 2021 und mögliche Stichprobe 2024), ob bspw. bei Studierenden aus vergleichbaren Gruppen Veränderungen der non-kognitiven Personenmerkmale zu beobachten sind. Dass im Zuge der Entwicklung von Bildungsangeboten immer auch die beiden Perspektiven ‚KI als Inhalt‘ und ‚KI als Werkzeug‘ mitgedacht werden müssen, wird schließlich durch die Beiträge aus dem dritten Bereich illustriert. Wie die Bildungsangebote ständig aktualisiert werden, stellt auch bei diesen beiden Beiträgen eine wesentliche Herausforderung dar.

Die Gastherausgeber*innen

Prof. Dr. Roland Happ, Universität Leipzig, Institut für Wirtschaftspädagogik, Universitätsprofessor für berufliche Bildung mit dem Schwerpunkt Wirtschaft

Prof. Dr. Kristina Kögler, Universität Stuttgart, Institut für Erziehungswissenschaft, Universitätsprofessorin für Berufspädagogik

Jun.-Prof. Dr. Jacqueline Schmidt, Technische Universität Dresden, Fakultät Wirtschaftswissenschaften, Juniorprofessorin für Wirtschaftspädagogik, insbesondere Digitalisierung in Bildungs- und Arbeitswelten

Dr. Marc Egloffstein, Universität Mannheim, Area Wirtschaftspädagogik, Wissenschaftlicher Mitarbeiter

Korrespondenz an: happ@wifa.uni-leipzig.de

Thementeil

Jacqueline Schmidt & Roland Happ

Validierung eines Tests zu Grundlagenwissen über Künstliche Intelligenz von angehenden Lehrkräften

Künstliche Intelligenz (KI) hat eine zunehmende Bedeutung für zahlreiche Lebens- und Arbeitsbereiche und wird als "Treiber der Digitalisierung" in der zweiten Welle der digitalen Transformation verortet. Die berufliche Bildung ist durch die KI-bedingten Veränderungen der Tätigkeitsprofile und daraus resultierender Kompetenzanforderungen an (zukünftige) Fachkräfte unmittelbar von der dynamischen Entwicklung betroffen. Für die Ausgestaltung der Qualifikationsprozesse ist das Lehrpersonal zuständig, in dessen Ausbildung KI-Inhalte bisher nicht verankert sind. Es kann davon ausgegangen werden, dass diese Gruppe über wenig Vorkenntnisse zu KI verfügt. Um belastbare Befunde über die Ausprägung des KI-Wissens von (angehenden) Lehrkräften zu generieren und darauf aufbauend adressatengerechte Aus- und Weiterbildungsangebote zu schaffen, werden Testinstrumente benötigt, mit denen das Wissen valide gemessen werden kann. Für neu entwickelte Instrumente sind in Anlehnung an die "Standards for Educational and Psychological Testing" der American Educational Research Association (AERA) et al. (2014) umfassende Analysen zur Beurteilung der Validität notwendig. Im Artikel werden die Analysen zur Beurteilung der Aspekte "Testinhalt" und "Beziehung zu anderen Merkmalen" diskutiert. Auf Basis der Befunde können die theoretisch postulierten Annahmen zum KI-Wissen von (angehenden) Lehrkräften mit empirischer Evidenz untersetzt werden.

Schlagworte: Digitale Transformation – Künstliche Intelligenz – Lehrkräfte – Validierung – Wissenstest

Validation of a test on prospective teachers' basic knowledge about artificial intelligence

Artificial intelligence (AI) is becoming increasingly important for numerous areas of life and work and is considered to be a "driver of digitalisation" in the second wave of digital transformation. Vocational education and training is directly affected by this dynamic development, given the AI-related changes to job profiles and the resulting skills requirements for (future) skilled workers. Responsibility for organising the qualification processes lies with the teaching staff, whose professional training so far lacks AI content. It can be assumed that this group has little prior knowledge of AI. In order to generate reliable findings on the extent of the AI knowledge of (prospective) teachers and to establish appropriate training and further education programmes based on this, test instruments are required that can be used to validly measure knowledge. For newly developed instruments, comprehensive analyses based on the "Standards for Educational and Psychological Testing" of American Educational Research Association (AERA) et al. (2014) are required to ensure validity. The article discusses the analyses regarding the assessment of the aspects "test content" and "relation to other characteristics". Based on the findings, the theoretically postulated assumptions regarding the AI knowledge of (prospective) teachers can be substantiated with empirical evidence.

Keywords: artificial intelligence – digital transformation – knowledge test – validation – vocational teachers

1 Relevance

The impact of digitalization on everyday life and work has been the subject of controversial debates over the past few decades (Schumann et al., 2022). Widespread use of digital technologies has led to changes in professional practices (Dubs, 2018; Geiser, 2022) and thus competence requirements for professionals in many fields (Geiser et al., 2021; Riebenbauer et al., 2022). Although a technology-oriented perspective often is taken in these discussions, a practical perspective is emerging regarding what knowledge and skills people need to survive in their dynamically changing world (Müller, 2018; Sczogiel et al., 2019) and how these competences can be fostered within educational processes. Vocational education plays a crucial role in creating a well-equipped workforce and thus is directly affected by this digital transformation (Winkler & Schwarz, 2021). Not only is the content of digital media as a subject in vocational education programs being discussed, but so is the potential of digital media in the design and delivery of those programs (Ständige Wissenschaftliche Kommission (SWK), 2022). Vocational teachers play a distinct role in digitization processes at two levels: they must understand and teach digital media content to students, and they must understand how to employ digital media to enhance teaching and learning processes (Gerholz et al., 2022; Meiners et al., 2022). The increasing importance of digital media in education has led to the establishment of several conceptual approaches to developing potential vocational teachers' digital competences (e. g., TPACK, Mishra & Koehler, 2006; DigCompEdu, Redecker, 2017). Over the past several years, the discussion has shifted towards the development and dissemination of artificial intelligence (AI), as it often has been characterized as the "driver of digitalization" (Federal Government, 2018, p. 10) initiating the second wave of digital transformation (Seufert et al., 2020; Winkler & Schwarz, 2021). In the context of educational processes, the widespread use of AI-based systems such as ChatGPT (OpenAI, 2023) has been controversial (Steppuhn, 2023). With continuous and rather easy access to such AI-based systems the question arises as to what knowledge and skills students and teachers need to be able to use them competently. Given the premise that teachers should be equipped with the same competences they seek to foster in their students (Guggemos & Seufert, 2020), it is essential that teachers' AI-related knowledge and competences be assessed and suitable training programs in AI be developed and implemented to meet their professional development needs. So far, no conceptual framework of AI-related knowledge that teachers need has been developed, and no instrument that takes into account crucial quality criteria of test theory has been created to assess that knowledge, both of which are essential for developing effective learner-centred training programs (Schmidt, 2024). Findings from current research (for an overview, see Schmidt, 2024) indicate a deficit in teachers' knowledge about AI (Hofstetter &

Massmann, 2021; Lindner & Berges, 2020; Lindner & Romeike, 2019; Nenner et al., 2021; Pfeiffer et al., 2021; Serholt et al., 2014), which seems to be mainly influenced by so-called 'hype topics' and their representation in the media (Lindner & Berges, 2020; Lindner & Romeike, 2019) and a lack of systematic and standardized procedures for test-based assessment of knowledge about AI. Further, in most of these studies focus is on (prospective) teachers in the general education sector; thus, prospective vocational teachers' knowledge of AI has been largely neglected in the literature (e. g., Roppertz, 2021).

Among other things, this led to the development of a structural model for the cognitive and non-cognitive AI-related competences of prospective vocational teachers in which knowledge is a central facet. A test based on this model was then designed to empirically assess the target group's knowledge about AI (Schmidt, 2024). To ensure valid test score interpretations, several analyses considering the Standards for Educational and Psychological Testing by the American Educational Research Association (AERA et al., 2014) have been conducted (Schmidt, 2024). According to the AERA et al. (2014) there are several sources of validity evidence. In this article, the evidence regarding the aspect test content and the test construct's relationship to external/other variables are discussed comprehensively as a preliminary step in the validation process. Once satisfactory findings are obtained on these two elements, other sources of validity evidence (i. e., response processes and consequences of testing) must be explored to obtain the differentiated data necessary to ensure validity of the test as a tool to assess (prospective) vocational teachers' knowledge about AI.

In this study, the following two questions concerning the validity of the assessment tool are addressed:

RQ1: To what extent does the content of this test allow valid interpretation of test scores?

RQ2: To what extent does the test construct's relations to other variables allow valid interpretation of test scores?

In the following, an overview is given of the validation process as set out in the Standards for Educational and Psychological Testing (AERA et al., 2014). Then, the procedures for designing and administering the test are outlined. Finally, the results are presented and ways in which these preliminary insights can be used to revise the instrument for further validation are discussed.

2 Validation Process

For the interpretation of test scores to be valid (Kane, 2013), an assessment instrument must be examined thoroughly for its validity. In theories on assessment, valid-

ity is considered the most important criterion for quality (e. g., Krohne & Hock, 2007; Rost, 2004). The validation process is intended to provide information about the extent to which the theoretical assumptions about the construct can be grounded with empirical evidence (Förster et al., 2017; Krohne & Hock, 2007; Meinhardt, 2018). Furthermore, the validation process requires analysis of conceptual assumptions (Krohne & Hock, 2007), which in this context refers to the substantive modeling of the measurement object. Thus, aggregation of several analyses allows an evaluation to be made of the extent to which the instrument allows valid interpretation of test values (Kane, 2013).

The Standards for Educational and Psychological Testing were established by the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) to serve as a guide particularly in empirical research in social and education sciences for developing and administering tests to ensure their validity (AERA et al., 2014; see, e. g., Kuhn, 2014; Happ, 2019; Schmidt, 2024). These standards address issues concerning test content, response processes, internal structure, relations to other variables, and consequences of testing (AERA et al., 2014).

To explore the extent to which the content of the test designed for this study reflects the construct under investigation accurately, the relevance, thoroughness, and representativeness of each test item were examined (Krohne & Hock, 2007). To explore the test construct's relations to other variables, the known groups method (Schnell et al., 2008) was employed and the test taker's field of study (contrast group I) and year of study in the same field (contrast group II) were taken into consideration.

In Table 1 an overview is given of the sources of validity explored in this paper and the methods employed to gather both qualitative and quantitative data.

Table 1: Sources of validity explored in the validation process of the test designed to assess prospective vocational teachers’ knowledge about AI (Schmidt, 2024)

Criteria examined for validity		Method
test content	1) Items as a representative set of the measurement object	Systematization of the measurement object (matrix)
	2) Relevance of the items in terms of content	Test development
	3) Adequacy of the items	
relations to other variables	4) Extent of knowledge about AI in a field of study	Method of known groups
	5) Extent of knowledge about AI in relation to the year of study	Method of known groups

3 Analyses

3.1 Test content

To determine the validity of the content of a test, the relationship between the conceptual considerations of the theoretical construct and its representation in test items must be examined (AERA et al., 2014). Accordingly, the extent to which the items on the test in this study adequately represented the content of the theoretical construct in question was explored (Krohne & Hock, 2007). The relevance and representativeness of the test items were of central importance in determining whether the test instrument adequately covered the measurement object in terms of both content and cognition. In Schmidt (2024), an extensive theoretical model was created to determine what basic knowledge about AI in terms of content and cognitive facets are needed to understand and apply basic concepts of AI. This model served as a basis for analyzing the validity of the test content, as it permitted consideration of the coherence between the theoretically modeled construct and its representation in the test items.

The content-related modeling was based on textbook analysis (N = 5)¹ and expert interviews (N = 4) (Meß, 2022). A horizontal perspective (Ernst, 2012) was adopted to analyze the textbooks. Due to the wide availability of online teaching-learning resources, online introductory works were considered in addition to print media.

¹ The analysis was aligned with the following works: Buxmann and Schmidt (2021); Deckert and Meyer (2020); Elements of AI (2022); KI-Campus (2020); Reinhart et al. (2021); Taulli (2019); Wittpahl (2019).

Because the target group was assumed to have limited knowledge about AI and due to a lack of curricular anchoring (Pargmann & Berding, 2022), some widely used standard works (e. g., Russel & Norvig, 2016) were excluded from analysis because they were designed for learners who had technical knowledge.² The experts interviewed were selected according to the criteria outlined by Gläser and Laudel (2010) and were chosen because they dealt with AI and the mediation of AI basics in the course of their daily work (e. g., in the context of consulting activities or through the development of learning opportunities on this topic). During the interviews the experts were questioned about the depth and scope of their knowledge about, and experience with, AI (see Bogner et al., 2014).

With these two qualitative approaches to collecting data the following four areas of knowledge of AI could be distinguished: basics of AI, methods of AI, applications of AI, and effects of AI. These four content areas are not to be considered in isolation; rather, they represent a more integrative content model (Meß, 2022) and formed the basis for specifying the measurement object of basic knowledge about AI in terms of test content.

Knowledge as a cognitive facet constitutes a fundamental element in models of teachers' professional competence (Seufert & Guggemos, 2022; Zlatkin-Troitschanskaia et al., 2013). It thus provides the foundation for all other cognitive performance dispositions (Happ, 2017; Minnameier, 2005). According to the understanding of Beck (1995) and Dubs (1995), basic knowledge about AI can be classified as declarative-static (knowledge of facts) and/or declarative-procedural (knowledge of structures) (Schmidt, 2024). By narrowing down the cognitive disposition of knowledge to knowledge of facts and knowledge of structures, restrictions must be imposed on the content-related modeling. Empirical findings from the modeling process indicate that effects of AI refer mainly to ethical-moral issues (Meß, 2022) that cannot be covered within the framework of the concept of knowledge specified in this study. For this reason, the three areas of basics of AI, methods of AI, and applications of AI were initially considered for the operationalization of knowledge of AI.³ The cognition-based modeling in this study was aligned with international models of AI literacy. The cognitive disposition of knowledge, which in this context includes remembering and understanding, forms the basis for all other cognitive performance dispositions (Ng et al., 2021).

² Although this widely used book essentially contains foundational information, it is aimed at computer scientists and computer science students. It was assumed that the students in the target group in this study did not have the technical knowledge that the book's target group is expected to have.

³ Further development of the model to include ethics/morality has been initiated (see Guggemos et al. in this issue).

Based on the content-related and cognition-related modeling, the assumptions made in chapter 2 regarding test content were evaluated. This applied especially to the relevance and representativeness of the test items. Furthermore, the appropriateness of the items for the intended target group had to be determined partly because validity must always be assessed in relation to the purpose of the test (Seufert et al., 2020). Within this framework, appropriateness was assessed by specifying the object of measurement as being knowledge aligned with the cognition-based modeling. This specification arose from the assumption that the target group has little to no knowledge about AI, which can be explained by the dynamic nature of the research topic and the fact that the participants had not received any formal training in AI as part of their university education. In addition, the participants likely had little practical teaching experience as they were in their first phase of teacher education; therefore, it could not be assumed that they had prior experience in using or even encountering AI in educational processes.

Based on the content- and cognitive-related modeling, a test consisting of 21 items in a single-choice format was developed (Schmidt, 2024; Schmidt & Happ, 2022). Considering the content-related modeling, seven items were designed to assess basics of AI, six to assess methods of AI, and eight to assess applications of AI (Schmidt, 2024). With reference to cognitive differentiation, 12 items were designed to assess declarative-static knowledge (knowledge of facts) and nine to assess declarative-procedural knowledge (knowledge of structures) (Schmidt, 2024). By considering both the content-related and cognition-related modeling, the validation criterion of test content could be considered fulfilled.

The connection between theoretical modeling and operationalization over the course of the test development is shown in the following matrix (see Table 2).

Table 2: Mapping of the Ttest items developed with respect to content-related and cognition-related differentiation (Schmidt, 2024)⁴

Content-related differentiation	Cognition-related differentiation	
	Declarative-static	Declarative-procedural
Basics of AI	W4, W8, W9, W10, W15, W20	W5
Methods of AI	W11, W12, W17	W2, W3, W13
Applications of AI	W7, W14, W19	W1, W6, W16, W18, W21

⁴ The items are included in Schmidt (2024), which is available as open access.

3.2 Relations to other variables

To evaluate a test construct's relations to other variables, either a convergent or a discriminant approach can be taken to conduct analyses (AERA et al., 2014). Convergent validation must be excluded in this study due to the lack of comparable test instruments (Schmidt, 2024). Discriminant validation, however, involves using the same instrument with different target groups. The known groups method (Schnell et al., 2008) can be employed to ground theoretically postulated differences in the manifestation of features using empirical evidence. For this study, a group of students in business informatics was considered a suitable contrast group. Due to the curricular structure of their study program, the students in contrast group I were expected to have greater knowledge about AI than the target group of prospective vocational teachers. Discriminant validation with regard to the test taker's stage of education (progress of study) also was conducted. The target group of prospective vocational teachers consisted of master degree students, and contrast group II consisted of bachelor degree students in the same field, which seemed suitable for comparison because it often is assumed that as one progresses through a course of study, one gains knowledge. However, neither the bachelor degree program nor the master degree program offered formal opportunities to learn about AI, which is why the level of knowledge about AI was not expected to differ significantly between these groups (Schmidt, 2024).

To obtain the quantitative data needed to examine the underlying assumptions, the test developed for this study was administered over the Limesurvey online platform (Limesurvey, 2011) and deployed at nine universities⁵ where prospective teachers in the vocational education sector are trained. For the purpose of discriminant validation, the test also was administered at two universities offering a business informatics program. Data collection took place between April 2021 and July 2023. The knowledge test was embedded in a larger assessment of cognitive and non-cognitive facets associated with AI (see Schmidt, 2024). In this study, however, the findings of the knowledge test were analyzed to determine the validity of the test items in terms of content and the construct's relations to other variables. As shown in Table 3, N = 191 prospective vocational teachers⁶ in master degree programs were

⁵ The test was administered in German to students at universities in Leipzig, Mainz, Hamburg, Kaiserslautern, Freiburg, Stuttgart, Göttingen, Schwäbisch Gmünd (Germany) and at a university in Zürich (Switzerland). Conducting the survey at numerous universities made it possible to cover not only the commercial/administrative area, but also the industrial/technical field and the area of healthcare, thus reflecting the three pillars of vocational and business education (Sektion Berufs- und Wirtschaftspädagogik, 2014) in the empirical dataset.

⁶ In this case, prospective teachers were all master degree students who could pursue a teaching career after completing a teacher traineeship.

surveyed as the target group, $N = 27$ students in a master degree program in business informatics⁷ were surveyed as contrast group I, and $N = 29$ prospective vocational teachers in a bachelor degree program in business and economics education were surveyed as contrast group II.⁸

Table 3: Sample

Group	Definition	N
Target group	Prospective vocational teachers in a master degree program	191
Contrast group I	Students in a master of business informatics program	27
Contrast group II	Prospective vocational teachers in a bachelor degree program in business and economics education	29

Notes: N = sample size

In the target group ($N = 191$), the subjects were on average 28.26 years old ($SD = 5.665$) and 56.5 % identified as female. Contrast group I subjects ($N = 27$) had a mean age of 23.7 years ($SD = 1.938$) and 37 % identified as female. The subjects of contrast group II ($N = 29$) were on average 22.5 years old ($SD = 4.702$) and 58.6 % identified as female. With regard to the fields of study, 71.2 % of the entire sample were in business and economics education, whereas 16.2 % were in the industrial-technical field, and 12.6 % were in the field of health and care.⁹

The quantitative data was analyzed using IBM SPSS software, version 27. For the calculations, the test scores were coded dichotomously: a correct answer on the knowledge test was given a score of 1, and an incorrect answer, regardless of which of the three incorrect answer alternatives was selected, was given a score of 0 in the data set (Schmidt, 2024). The scores were tallied so that a maximum of 21 points could be achieved on the knowledge test.

⁷ To assess this contrast group, the test was administered in German at Leipzig University and in English at the University of Wrocław (English version of the test).

⁸ The second contrast group was assessed at Leipzig University.

⁹ In the data set no significant differences were found in the relevant variables between the subgroups, meaning that the target group of prospective vocational teachers relevant for the analyses included all of these groups.

Table 4: Average test performance of the target group and contrast groups I and II

	N	M	Correct response rate	SD
Target group	191	8.90	0.424	2.61
Contrast group I (business informatics)	27	10.96	0.522	3.93
Contrast group II (bachelor students of business and economics education)	29	8.66	0.412	2.30

Notes: N = sample size, M = mean, SD = standard deviation

The findings indicate that on average, the target group solved fewer than half of the test items correctly (Table 4), scoring 8.90 (SD = 2.61) points on average, which equates to a correct response rate of 42.4 %. Contrast group I performed better, with an average correct response rate of 52.2 %, totalling 10.96 (SD = 3.927) points. The mean difference was significant on the basis of inferential statistical analyses ($p < .01$)¹⁰. According to Cohens' d (0.735), this effect can be considered moderate. The test values of these two groups also reveal with regard to the standard deviation that the extent of knowledge about AI ranged more widely among business informatics students than it did among prospective vocational teachers. Contrast group II achieved an average of 8.66 (SD = 2.30) points on the knowledge test. This value was only marginally below the average value of the target group. Thus, there was no significant mean difference between the target group and contrast group of prospective teachers. This is further supported by Cohens' d (-0.093).

More differentiated conclusions about the level of knowledge of the respective groups can be drawn by analyzing the response patterns with regard to the theoretical content-related and cognition-related internal differentiation of the construct. The average correct response rate of the item sets for each content area are presented in the following (see Table 5).

¹⁰ Results of the Levene test of equality of variance showed no homoscedasticity could be assumed for the two groups. Although the t-test for independent samples is fairly robust to any violation of the prerequisites, an increased error rate must be assumed for samples of significantly different sizes and variance inequality (Bortz, 2005). For this reason, non-parametric tests were used in this case and the general tendency was determined using the Mann-Whitney test based on test scores.

Table 5: Correct response rates of each group with regard to content areas

	Basics of AI	Methods of AI	Applications of AI
Target group	0.473 (0.209)	0.462 (0.218)	0.349 (0.156)
Contrast group I (business informatics)	0.640 (0.249)	0.548 (0.244)	0.407 (0.165)
Contrast group II (bachelor degree program in business and economics education)	0.429 (0.162)	0.466 (0.191)	0.358 (0.166)

The results indicate that the target group did better on items related to basics of AI with an average correct response rate of 0.473 than on items related to methods of AI with a rate of 0.462 and on items related to applications of AI with 0.349. This reveals that the lowest number of items could be solved correctly in the area "Applications of AI" among the target group.

Predictably, contrast group I (students of business informatics) outperformed the target group in all content areas. Moreover, unlike the other two groups, contrast group I solved more than 50 % of all the items correctly. With an average correct response rate of 0.640, contrast group I solved almost two thirds of the items related to basics of AI correctly. This area yielded the highest average correct response rate for this group. Contrast group II (bachelor degree students in business and economics education) performed similarly to the target group, which was to be expected from the global analysis of the test scores. What is noticeable at this point is that the content area with the highest average correct response rate was methods of AI.

In addition to the analysis of individual content areas, an analysis according to the cognition-related differentiation of the measurement object was conducted. Based on the theoretical modeling, the items were assigned to declarative-static knowledge and declarative-procedural knowledge. The average correct response rates of the subsets can be found in Table 6.

Table 6: Correct response rates of each groups with regard to the cognition-related modeling

	Declarative-static		Declarative-procedural	
Target group	0.381	(0.144)	0.479	(0.163)
Contrast group I (business informatics)	0.512	(0.168)	0.535	(0.249)
Contrast group II (business and economics education)	0.365	(0.134)	0.475	(0.175)

The findings reveal that all groups performed better in the area of declarative-procedural knowledge (knowledge of structures) than in the area of declarative-static knowledge (knowledge of facts). It is striking that the differences in the average correct response rates between these two facets were only marginally distinct within the group of business informatics students, whereas these differences were more prominent within the target group and within contrast group II.

According to these findings, the assumptions derived in chapter 2 regarding the source of validity evidence relations to other variables can be confirmed.

4 Discussion, limitations, and further steps

The aim of this article was to determine the validity of scores on a test instrument developed for assessing prospective vocational teachers' knowledge of AI. The test was aligned with the Standards for Educational and Psychological Testing (AERA et al., 2014) and the sources of validity evidence under investigation were test content and relations to other variables. Both qualitative and quantitative approaches were taken to gathering data.

According to the results, the measurement object modeled was represented accurately in the test instrument, taking into account the content areas that emerged from the textbook analyses and expert interviews (Meß, 2022). This conclusion is drawn from the roughly equal distribution of items over the three content areas basics of AI, methods of AI, and applications of AI. By taking these content areas into account, the relevance of the test items were ensured.

In addition, cognition-based modeling of declarative-static knowledge (knowledge of facts) and declarative-procedural knowledge (knowledge of structures) was also considered when developing the test. The almost equal division of the test items into these two categories of knowledge ensured adequacy of the items. Thus, the assumptions established were sufficiently tested within the scope of the validation of test content when answering RQ1.

Quantitative analyses were conducted to test the assumptions in the context of the validation aspect "relations to other variables". For this purpose, two contrast groups were assessed using a discriminant approach and subsequently analyzed with respect to their mean test performance. The findings indicate that, in line with theoretically derived expectations, the students of business informatics (contrast group I) performed better on the test than the target group, confirming the assumption that this contrast group has more advanced knowledge about AI than the target group. Furthermore, the assumption was made that the target group's knowledge about AI did not differ significantly from the AI knowledge of bachelor degree students in the

same subject area. Based on the quantitative findings, this assumption was verified. By confirming this assumption, RQ2 was answered.

To improve the test instrument developed in this study, further analyses are necessary. According to the Standards for Educational and Psychological Testing (AERA et al., 2014), response processes and consequences of testing are other important sources of validity evidence and must be explored. Furthermore, a follow-up interview with experts would be beneficial to assess the relevance of the test items developed (Beck, 2020). The aim of future research will be to adjust the instrument more specifically to the intended target group and thereby increase the validity of test score interpretations. The aim of the analyses conducted in this study was to assess the validity of test score interpretations based on conceptual considerations and empirical findings. The findings of this study will improve the further development of the test so that it will be more effective in assessing the knowledge of the target group in question. Particular attention will be paid to factors influencing knowledge, the development of knowledge over time, and possible reciprocal effects on other facets.

In this study, only prospective vocational teachers' knowledge of AI was explored; in-service teachers' knowledge of AI will be investigated in future research, as this information is essential for developing effective, research-based professional development programs designed to meet the particular needs of the target group. Addressee-appropriate learning opportunities have great potential, as results of a recent intervention study indicate that (prospective) teachers were able to increase their AI knowledge significantly within the time span of only one semester (Schmidt, 2024; Schmidt & Happ, 2022). The approach has already been initiated in a large-scale collaborative project (KIWi-MOOC) within the framework of the WÖRLD Competence Center (Thiel de Gafenco & Klusmeyer, 2023).

References

- American Educational Research Association (AERA), American Psychological Association (APA) & National Council of Measurement in Education (NCME) (2014). Standards for educational and psychological testing. Washington: American Educational Research Association.
- Beck, K. (1995). Wirtschaftskundliches Wissen bei kaufmännischen Lehrlingen – Zur Diagnose kognitiver Strukturen mit dem WBT [Business knowledge in commercial apprentices – On the diagnosis of cognitive structures with the WBT]. In C. Metzger & H. Seitz (Hrsg.), *Wirtschaftliche Bildung: Träger, Inhalte und Prozesse* (S. 375-395). Zürich: Verlag des Schweizerischen Kaufmännischen Verbandes.
- Beck, K. (2020). Ensuring content validity of psychological and educational tests – The role of experts. *Frontline Learning Research*, 8 (6), 1-37.
- Bogner, A., Littig, B. & Menz, W. (2014). Interviews mit Experten. Eine praxisorientierte Einführung [Interviews with experts: A practice-oriented introduction]. Wiesbaden: Springer.
- Bortz, J. (2005). Statistik für Human- und Sozialwissenschaftler [Statistics for human and social scientists] (6. Aufl.). Berlin: Springer.

- Buxmann, P. & Schmidt, H. (2021). Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. [Artificial intelligence. With algorithms to economic success] (2. Aufl.). Berlin: Springer Gabler.
- Deckert, R. & Meyer, E. (2020). Digitalisierung und Künstliche Intelligenz. Kooperation von Menschen und Maschinen aktiv gestalten [Digitalization and artificial intelligence. Actively shaping cooperation between people and machines]. Wiesbaden: Springer Fachmedien.
- Dubs, R. (1995). Lehrerverhalten: Ein Beitrag zur Interaktion von Lehrenden und Lernenden im Unterricht [Teacher behavior: A contribution to teacher-learner interaction in the classroom]. Zürich: Verlag des Schweizerischen Kaufmännischen Verbandes.
- Dubs, R. (2018). Gedanken zur Zukunft der kaufmännischen Grundbildung [Thoughts on the future of basic commercial education]. In J. Schlicht & U. Moschner (Hrsg.), Berufliche Bildung an der Grenze zwischen Wirtschaft und Pädagogik: Reflexionen aus Theorie und Praxis (S. 37-58). Wiesbaden: Springer VS.
- Elements of AI. (2022). Einführung in die KI. [Introduction to AI]. Verfügbar unter: <https://course.elements-ofai.com/de/> [10.01.2024].
- Ernst, F. (2012). Fachdidaktische Analyse von Lehrbüchern für den Rechnungswesenunterricht in Deutschland und den USA [Didactic analysis of accounting textbooks in Germany and the U.S.]. Dissertation. Friedrich-Schiller-Universität Jena.
- Federal Government (2018). Artificial intelligence strategy. Verfügbar unter: https://www.ki-strategie-deutschland.de/home.html?file=files/downloads/Nationale_KI-Strategie_engl.pdf&cid=729 [15.11.2018].
- Förster, M., Happ, R. & Molerov, D. (2017). Using the U. S. test of financial literacy in Germany – Adaptation and validation. *Journal of Economic Education*, 48 (2), 123-135. <https://www.doi.org/10.1080/00220485.2017.1285737>
- Geiser, P. (2022). Lehrerüberzeugungen zur Bedeutung der Digitalisierung: Eine Interviewstudie mit Lehrkräften zur Ausbildung kaufmännischer Fachkräfte [Teachers' beliefs about the importance of digitization. An interview study with teachers training commercial specialists]. Bielefeld: wbv.
- Geiser, P., Busse, J., Seeber, S., Schumann, S., Weber, S., Zarnow, S., Hiller, F., Hackenberg, T. & Lange, A. (2021). Kompetenzen in digitalisierten kaufmännischen Arbeitsplatzsituationen [Competencies in digitalized commercial workplace situations]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117 (4), 630-657.
- Gerholz, K.-H., Schlottmann, P., Slepcevic-Zach, P. & Stock, M. (2022). Digital Literacy in der beruflichen Lehrerinnen- und Lehrerbildung – Einleitung [Digital literacy in professional teacher education – Introduction]. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation (S. 11-18). Bielefeld: wbv.
- Gläser, J. & Laudel, G. (2010). Experteninterviews und qualitative Inhaltsanalyse [Expert interviews and qualitative content analysis]. Springer VS.
- Guggemos, J. & Seufert, S. (2020). Teaching with and teaching about technology – Evidence for professional development of in-service teachers. *Computers in Human Behavior*, 115, 1-11.
- Happ, R. (2017). Die Entwicklung des volkswirtschaftlichen Grundlagenwissens im Studienverlauf – Effekte von Eingangsvoraussetzungen auf den Wissenserwerb [The development of basic economic knowledge in the course of study – Effects of entry requirements on knowledge acquisition]. Landau: Verlag Empirische Pädagogik.
- Happ, R. (2019). Der Einfluss des Migrationshintergrundes auf unterschiedliche Facetten des ökonomischen Wissens – Implikationen für die Wirtschaftsdidaktik [The influence of migration background on different facets of economic knowledge – Implications for business education]. Kumulative Habilitationsschrift. Mainz: Johannes-Gutenberg-Universität.
- Hofstetter, A. & Massmann, C. (2021). Digital skills gap report 2020 – Skills we need, skills we teach. Verfügbar unter: <https://kohorten.com/2020/05/13/digital-skill-gap-studie/> [05.08.2023].
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50 (1), 1-73.
- KI-Campus (2020). Einführung in die KI [Introduction to AI]. Verfügbar unter: <https://ki-campus.org/courses/einfuehrungki2020> [05.05.2023].

- Krohne, H.-W. & Hock, M. (2007). *Psychologische Diagnostik: Grundlagen und Anwendungsfelder* [Psychological diagnostics: Fundamentals and fields of application]. Stuttgart: Kohlhammer.
- Kuhn, C. (2014). *Fachdidaktisches Wissen von Lehrkräften im kaufmännisch-verwaltenden Bereich. Modellbasierte Testentwicklung und Validierung* [Subject didactic knowledge of teachers in the commercial-administrative field. Model-based test development and validation]. Landau: Verlag Empirische Pädagogik.
- Limesurvey. (2011). LimeSurvey: The open source survey application.
- Lindner, A. & Berges, M.-P. (2020). Can you explain AI to me? Teachers' pre-concepts about artificial intelligence. In 2020 IEEE Frontiers in Education Conference (FIE). Uppsala, SE.
- Lindner, A. & Romeike, R. (2019). Teachers' perspectives on artificial intelligence. Conference Paper. ISSEP 2019. 12th International conference on informatics in schools: Situation, evaluation and perspectives.
- Meiners, H., Hartmann, P., Niegemann, H., Seeber, S., Wuttke, E. & Schumann, S. (2022). Digitale Medienkompetenz als Voraussetzung für die Erstellung von Prüfungsaufgaben [Digital media literacy as a prerequisite for the creation of examination tasks]. In S. Schumann, S. Seeber & S. Abele (Hrsg.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (S. 123-144). Bielefeld: wbv.
- Meinhardt, C. (2018). Entwicklung und Validierung eines Testinstruments zu Selbstwirksamkeitserwartungen von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern [Development and validation of a test instrument on self-efficacy expectations of (prospective) physics teachers in physics didactic fields of action]. Berlin: Logos Verlag. <https://www.doi.org/10.30819/4712>
- Meß, M. (2022). Wissenstests zu Anwendungen der Künstlichen Intelligenz – Validierung von Items eines Fragebogens durch Experteninterviews [Knowledge tests on artificial intelligence applications – Validation of items of a questionnaire by expert interviews]. Masterarbeit. Universität Leipzig.
- Minnameier, G. (2005). *Wissen und inferentielles Denken* [Knowledge and inferential thinking]. Berlin: Peter Lang.
- Mishra, P. & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108 (6), 1017-1054.
- Müller, (2018). Berufsbildung 4.0. Was müssen die Fachkräfte der Zukunft können? [Vocational training 4.0 - what do the skilled workers of the future need to master?] In C. Bär, T. Grädler & R. Mayer (Hrsg.), *Digitalisierung im Spannungsfeld von Politik, Wirtschaft, Wissenschaft und Recht* (S. 292 – 298). Berlin, Heidelberg: Springer.
- Nenner, C., Damnik, G. & Bergner, N. (2021). Weil informatorische Bildung nicht erst in der Sekundarstufe beginnen darf: Integration informatorischer Bildung ins Grundschullehrerstudium [Because computer literacy should not start in secondary school: Integrating information technology education into elementary school teacher education]. In L. Humbert (Hrsg.), *Informatik – Bildung von Lehrkräften in allen Phasen* (S. 103-112). Gesellschaft für Informatik.
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Shen, M. Q. (2021). Conceptualizing AI literacy: An exploratory overview. *Computers and Education: Artificial Intelligence*, 2, 1-11. <https://www.doi.org/10.1016/j.caeai.2021.100041>
- OpenAI. (2023). GPT-4 Technical Report. <https://www.doi.org/10.48550/arXiv.2303.08774>
- Pargmann, J. & Berding, F. (2022). Künstliche Intelligenz in der beruflichen Bildung – Utopie oder Dystopie? Chancen und Herausforderungen von KI als Bildungstechnologie [Artificial intelligence in vocational education – Utopia or dystopia? Opportunities and challenges of AI as an educational technology]. In M. Eckelt, T. Ketschau, J. Klassen, J. Schauer, J. Schmees & C. Steib (Hrsg.), *Berufsbildungspolitik: Strukturen – Krise – Perspektiven* (S. 215-232). <https://www.doi.org/10.3278/9783763972623>
- Pfeiffer, A., Bezzina, S., Dingli, A., Wernbacher, T., Denk, N. & Fleischhacker, M. (2021). Adaptive learning and assessment: From the teachers' perspective. *Proceedings of INTED 2021 Conference*.
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. Publications Office of the European Union. Verfügbar unter: <https://publications.jrc.ec.europa.eu/repository/handle/JRC107466> [05.08.2023].

- Reinhart, J., Mayer, O. & Greiner, C. (2021). Künstliche Intelligenz. Grundlagen, Anwendungsbeispiele und Umsatzstrategien für Unternehmen [Artificial intelligence. Basics, application examples and revenue strategies for companies]. Würzburg: Vogel Buchverlag.
- Riebenbauer, E., Berding, F. & Flick-Holtsch, D. (2022). Fachwissenschaftliche Implikationen für die Aus- und Fortbildung von Lehrkräften zur Digitalisierung im Rechnungswesen [Subject-specific implications for teacher education and training on digitization in accounting]. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation. (S. 199-212). Bielefeld: wbv.
- Roppertz, S. (2021). Die Rolle und Bedeutung von Künstlicher Intelligenz in der Berufsausbildung – Implikationen für angehende Berufs- und Wirtschaftspädagogog*innen [The role and importance of artificial intelligence in vocational education – Implications for prospective vocational and business educators]. bwp@, 40, 1-23.
- Rost, J. (2004). Lehrbuch Testtheorie – Testkonstruktion (2. Vollst. überarb. und erw. Aufl.) [Textbook Test Theory - Test Construction]. Bern: Verlag Hans Huber.
- Russel, S. J. & Norvig, P. (2016). Artificial intelligence: A modern approach. Upper Saddle River: Prentice Hall.
- Schmidt, J. (2024). Grundlagenwissen zu Künstlicher Intelligenz von angehenden Lehrkräften – Modellbasierte Testentwicklung und Validierung [Basic knowledge about artificial intelligence of (prospective) teachers – Model-based test development and validation]. Bielefeld: Wbv. <https://www.doi.org/10.3278/9783763976539>
- Schmidt, J. & Happ, R. (2022). Fachspezifisches Wissen, Einstellungen und Überzeugungen zu Anwendungen der Künstlichen Intelligenz bei Studierenden der Wirtschaftspädagogik [Subject-specific knowledge, attitudes and beliefs about applications of artificial intelligence among students of business and economics education]. In K. Kögler, U. Weyland & H. H. Kremer (Hrsg.), Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2022 (S. 89-102). Leverkusen: Verlag Barbara Budrich.
- Schnell, R., Hill, P. B. & Esser, E. (2008). Methoden der empirischen Sozialforschung [Methods of empirical social research] (8. Aufl.). München: Oldenbourg.
- Schumann, S., Seeber, S. & Abele, S. (Hrsg.). (2022). Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen. [Digital transformation in vocational education: concepts, findings and challenges] Bielefeld: wbv.
- Sczogiel, S., Malapally, A., Schmitt-Rueth, S. & Williger, B. (2019). Future Digital Job Skills: Die Zukunft kaufmännischer Berufe [Future digital job skills: The future of commercial professions]. Industrie- und Handelskammer Nürnberg für Mittelfranken.
- Sektion Berufs- und Wirtschaftspädagogik [BWP] (2014). Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik im Rahmen berufs- und wirtschaftspädagogischer Studiengänge [Basic curriculum for the university subject of vocational and business education as part of vocational and business education degree programmes]. Verfügbar unter: https://www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf [16.08.2023].
- Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., Vasalou, A. & Castellano, G. (2014). Teachers' views on the use of empathic robotic tutors in the classroom. The 23rd IEEE International Symposium on Robot and Human Interactive Communication, 955-960.
- Seufert, S. & Guggemos, J. (2022). Digitale Kompetenzen von Lehrpersonen – Ergebnisse einer empirischen Untersuchung in der Berufsbildung [Digital competencies of teachers – Results of an empirical study in vocational education]. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach & M. Stock (Hrsg.), Digital Literacy in der beruflichen Lehrer:innenbildung. Didaktik, Empirie und Innovation. (S. 213-226). Bielefeld: wbv.
- Seufert, S., Guggemos, J., & Sailer, M. (2020). Technology-related knowledge, skills and attitudes of pre-service and in-service teachers: The current situation and emerging trends. Computers in Human Behavior, 115, 1-7. <https://www.doi.org/10.1016/j.chb.2020.106552>

- Ständige Wissenschaftliche Kommission (SWK). (2022). Digitalisierung im Bildungssystem: Handlungsempfehlungen von der Kita bis zur Hochschule: Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK) [Digitization in the education system: Recommendations for action from daycare to higher education: Report by the standing scientific commission of the standing conference of the ministers of education and cultural affairs of the Länder in the federal republic of Germany (SWK)]. Verfügbar unter: https://www.kmk.org/fileadmin/Dateien/pdf/KMK/SWK/2022/SWK-2022-Gutachten_Digitalisierung.pdf [05.08.2023].
- Steppuhn, D. (2023). Der Einsatz Künstlicher Intelligenz in Schulen [The use of artificial intelligence in schools]. *Bildung und Beruf*, 6, 122-125.
- Taulli, T. (2019). Artificial intelligence basics: A non-technical introduction. New York: Apress.
- Thiel de Gafenco, M. & Klusmeyer, J. (2023). Akteurs-Netzwerk ‚WÖRLD-House‘: Digitalisierungsbezogene Kompetenzentwicklung in der wirtschaftswissenschaftlichen Lehrkräftebildung [Actor network 'WÖRLD-House': Digitization-related competence development in economics teacher education]. Posterbeitrag: Zukunftswerkstatt Lehrkräftebildung Neu Denken des Stifterverbandes, Berlin, 25. Mai 2023. Verfügbar unter: <https://www.stifterverband.org/download/file/fid/12307> [22.08.2023].
- Winkler, F. & Schwarz, H. (2021). IT-Berufe im Wandel [IT professions in transition]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, Beiheft 31: Künstliche Intelligenz in der Berufsbildung, 77-102.
- Wittpahl, V. (Hrsg.). (2019). Künstliche Intelligenz. Technologie, Anwendung, Gesellschaft [Artificial intelligence. Technology, application, society]. Berlin: Springer Verlag. <https://www.doi.org/10.1007/978-3-662-58042-4>
- Zlatkin-Troitschanskaia, O., Happ, R., Förster, M., Preuße, D., Schmidt, S. & Kuhn, C. (2013). Analyse der Ausprägung und Entwicklung der Fachkompetenz von Studierenden der Wirtschaftswissenschaften und der Wirtschaftspädagogik [Analysis of the expression and development of professional competence of students of economics and business education]. In O. Zlatkin-Troitschanskaia, R. Nickolaus & K. Beck (Hrsg.), *Kompetenzmodellierung und Kompetenzmessung bei Studierenden der Wirtschaftswissenschaften und der Ingenieurwissenschaften* (S. 69-92). *Lehrerbildung auf dem Prüfstand (Sonderheft)*. Landau: Verlag Empirische Pädagogik.

Autor*innen

Jun.-Prof. Dr. Jacqueline Schmidt, Technische Universität Dresden, Fakultät Wirtschaftswissenschaften, Juniorprofessur für Wirtschaftspädagogik, insb. Digitalisierung in Bildungs- und Arbeitswelten
 Prof. Dr. Roland Happ, Universität Leipzig, Institut für Wirtschaftspädagogik, Professur für Berufliche Bildung mit dem Schwerpunkt Wirtschaft
 Korrespondenz an: jacqueline.schmidt@tu-dresden.de

Jule Hangen & Eveline Wuttke

Künstliche Intelligenz verstehen lernen – Ergebnisse eines Trainingsprogramms zur Förderung von Startup-Fähigkeiten im Bereich KI

Künstlicher Intelligenz (KI) wird als Schlüsseltechnologie ein hohes Potential für u. a. Produktivitätszuwächse zugeschrieben; daher werden Start-ups, die KI-Innovationen anbieten, als entscheidend für Wirtschaftswachstum angesehen. Allerdings scheitern viele neue Unternehmen in den ersten Jahren, und die Wirksamkeit von Trainingsprogrammen für (angehende) Unternehmer*innen wird selten systematisch evaluiert. In dieser Studie untersuchen wir das Wissen, die Motivation, die Überzeugungen und die Einstellungen von potenziellen Gründer*innen im Rahmen der Teilnahme an einem Trainingsprogramm. Dazu wurden Fragebögen und Interviews eingesetzt. Die Interviews zeigten vielfältige Gründe für die Teilnahme am Programm auf, u. a. die Suche nach Mitgründer*innen und das Interesse an KI. Darüber hinaus schätzten die Teilnehmer*innen den Nutzen und die Relevanz von KI für ihre Arbeit höher ein als die Benutzerfreundlichkeit. Die Evaluationsergebnisse zeigen keine signifikante Wissensveränderung über Anwendungen der KI, aber das selbstberichtete Wissen über Unternehmensgründungen nimmt zu. Die intrinsische Motivation war nach dem Programm signifikant niedriger. Das Programm kann insgesamt als effektiv hinsichtlich der Vermittlung von Wissen über Unternehmensgründungen gesehen werden, in Zukunft sollte jedoch der KI Fokus noch stärker betont werden, um die Gründung von KI oder data-driven ventures zu unterstützen.

Schlagwörter: Entrepreneurship Education – Gründungen – Künstliche Intelligenz – Trainingsprogramm

Learning about artificial intelligence – results from a training program to promote startup skills in the AI field

Artificial intelligence (AI) is key to productivity growth in the 21st century; thus, startups offering AI innovations are considered vital for the economy. However, many new businesses fail in their first few years, and the effectiveness of entrepreneurial training programs seldom is evaluated systematically. In this study, we investigate the knowledge, motivation, beliefs, and attitudes of potential founders of AI startups upon starting and completing a training program we designed for such potential entrepreneurs. We administered a questionnaire and conducted interviews with people enrolled in the program. We found participants had varying and often multiple reasons for enrolling in the training program, for example, to find business partners and get a motivational boost, and they rated the usefulness and job relevance of AI greater than its ease of use. Although we found no significant change in participants' knowledge about AI, we noticed a substantial increase in their knowledge about starting a business. Interestingly, their intrinsic motivation was significantly lower after the program. Overall, our training program for AI startups was effective in providing general startup knowledge but needs to emphasize the AI focus to support creation of AI or data-driven ventures.

Keywords: artificial intelligence – entrepreneurship education – startups – training study

1 Introduction

Artificial intelligence (AI) is considered key to productivity growth in the 21st century (Bundesregierung, 2020). The economic importance of AI for Germany becomes clear when considering that an estimated 220.6 billion euros in domestic annual sales are influenced by AI applications (Brandt, 2019). Although startups in the field of AI, particularly those involving the transfer of knowledge about AI technologies from research to practical application, are considered vital for the economy, many new ventures fail within their first few years. According to Plümer and Niemann (2016) approximately 50 % of such failures (not only in the domain of AI) are due to a lack of entrepreneurial skills, and Wang, Yang, Han, Huang and Wu (2022) find that 95 % are due to developmental problems during the postformation phase when entrepreneurs have to manage economic crises (see also Heinrichs, 2016, 2021).

It has been argued that for AI or data-driven startups to be successful, innovative “ecosystems” of AI talent are needed to generate ideas, transform knowledge into practical applications, and encourage and retain talents (Ecker, Coulon & Kohler, 2021). Ideally, such an ecosystem would involve innovators, investors, and institutions working together to meet the challenge of melding ideas for AI-based startups with knowledge about AI technologies to implement them in AI- or data-driven business models. Founders of such startups are expected to have knowledge about AI, motivation and positive beliefs about AI as well as a positive attitude toward AI. They furthermore need basic entrepreneurial knowledge to launch the new venture and sustain its development thereafter (Alberti, 1999). Entrepreneurial knowledge encompasses understanding of business concepts and skills, and it is key to a successful launch of a new venture and its sustained success (Wu, Chang & Chen, 2008). Through entrepreneurial training with a focus on knowledge, motivation and beliefs, one can develop the competences necessary to take on the challenges of establishing a business (Maresch, Harms, Kailer & Wimmer-Wurm, 2016), launching it, and developing it during the post-founding phase. Effective trainings, following research based standards for trainings, address the aforementioned qualification issues and provide new perspectives in various areas – in this case AI-related and data-driven new ventures. When designing entrepreneurial training programs, established criteria for successful trainings (Cademartori et al., 2017) should be considered and the impact of the programs should be assessed to determine their effectiveness. To this end, participants’ perceptions, learning gains, and behavioral responses should be investigated (Kirkpatrick & Kirkpatrick, 2006, for more details on design criteria and evaluation levels see chapter 2.3). The effectiveness of entrepreneurial training programs in general, and startup training programs in particular, often is not evaluated systematically (Nabi, Holden & Walmsley, 2006).

We developed, implemented and systematically evaluated a startup training in the domain of data-driven businesses. Participants intend to form a startup in the field of data-driven ventures. Knowledge about AI applications or knowledge about AI models were part of the program content. In each subsequent edition, new developments (such as ChatGPT) as well as traditional machine learning solutions were considered. The general research question is, if the training is able to promote the AI related entrepreneurial competence (knowledge, attitudes, beliefs and motivation) of potential founders in AI or data driven businesses.

In the following chapters we first explore the theoretical background of entrepreneurship education, and we describe the training program we designed accordingly for potential founders of AI startups. Next, we present our research questions and explain our method of data collection to determine the effectiveness of the program. Finally, we present our findings and discuss their implications.

2 Theoretical background

2.1 Relevance and effects of entrepreneurship education

Entrepreneurship education has become increasingly popular over the past few decades (Sreenivasan & Suresh, 2023) with the growing awareness that the economy depends to a certain degree on new ventures to stimulate economic growth (Moberg, 2014). Technologically innovative startups contribute to progress through market novelties and thus drive structural change (KfW, 2015). Entrepreneurship is a multifaceted endeavor, and the success of a new venture is affected by a variety of social, cultural, environmental, demographic and economic factors (Gaddam, 2007). Stamboulis and Barlas (2014) list economic, psychological (personal characteristics of the entrepreneur), social (e. g., consumer habits), environmental (e. g., access to resources, and economic climate), demographic, and cultural (e. g., shared values and beliefs) factors that influence the success of a new business. Some factors might hinder potential entrepreneurs from committing themselves to entrepreneurial ventures. These include negative beliefs about entrepreneurship (e. g., status and risk) and a lack of competence, experience, and basic knowledge about business.

Against this background, entrepreneurship education is deemed crucial and thus is already being promoted and integrated into school curricula in many European countries (European Commission, 2006) and in the United States (Kuratko, 2005). A key assumption underlying programs is that entrepreneurial skills can be taught and the success of a startup is not dependent on personal characteristics (Oosterbeek, van Praag & Ijsselstein, 2010); therefore, it is expected that through entrepreneurship education individuals will develop the competences necessary to take on the

challenges of establishing and running a business (Maresch et al., 2016; Roxas, Cayoca-Panizales & Mae de Jesus., 2008).

In the literature, findings concerning the impact of entrepreneurial training programs are scarce, and views sometimes are divergent. Of the few investigations into the effectiveness of such programs, most have been exploratory and descriptive and have yielded inconsistent findings. In particular, the role of training programs and their impact on entrepreneurial behavior remains unclear (Roxas et al., 2008).

2.2 Importance of knowledge, motivation, beliefs, and attitudes regarding entrepreneurship in general and AI-related entrepreneurship in particular

The success of learning processes has long been defined by output (e. g., Short, 1985; Klieme et al., 2003). The focus is in general not only on the acquisition of knowledge, but rather on the acquisition of competences (Norris, 1991) understood in terms of dispositions, usually encompassing knowledge, skills, and attitudes, and expressed in terms of coping with (occupational) requirements through the use of those dispositions (Hoffmann, 1999; Hartig, 2008; Weinert, 2001; Zlatkin-Troitschanskaia & Seidel, 2011). The use of competences in observable action is further influenced by motivational, volitional, and social factors. Among the most agreed upon definitions of competence in the German-speaking academic world is that of Weinert (2001), who describes it as the cognitive abilities and skills available to an individual or that can be learned by an individual to solve specific problems, as well as the motivational, volitional, and social readiness associated with them so that solutions to problems can be used successfully and responsibly in variable situations.

In entrepreneurship education, it often is assumed that entrepreneurial knowledge and access to it are a learner's most important resources (Widding, 2005). Although this popular view suggests that entrepreneurship education influences behavior and intentions to start a business, the causal relationship remains unclear. This could be seen as an indication that developing entrepreneurial knowledge is important for starting a business but that it is not enough. Beliefs, attitudes, and self-efficacy might play an important role as well (Roxas et al., 2008). Attitudes are described as a tendency resulting from the degree of approval or disapproval of something (Bohner, 2002). Expressions of attitude are influenced in particular by cognitive and affective dispositions (Siegfried & Wuttke, 2021) and exert an influence on intentions to act (Ajzen, 1991). Maresch et al. (2016) argue for example, that when one evaluates the outcome of starting a business as positive, one will have a more favorable attitude toward starting a business and consequently greater intention to

do so. They also found that pro-entrepreneurial attitudes correlated positively with entrepreneurial intentions.

Following the framework of Schmidt and Happ (2022b), the facets of entrepreneurial competence mentioned above which are deemed necessary to launch an AI startup successfully, are as follows (see Figure 1):

- ▶ knowledge about AI applications;
- ▶ a positive attitude toward AI and positive beliefs about AI;
- ▶ extrinsic and intrinsic motivation toward AI applications.

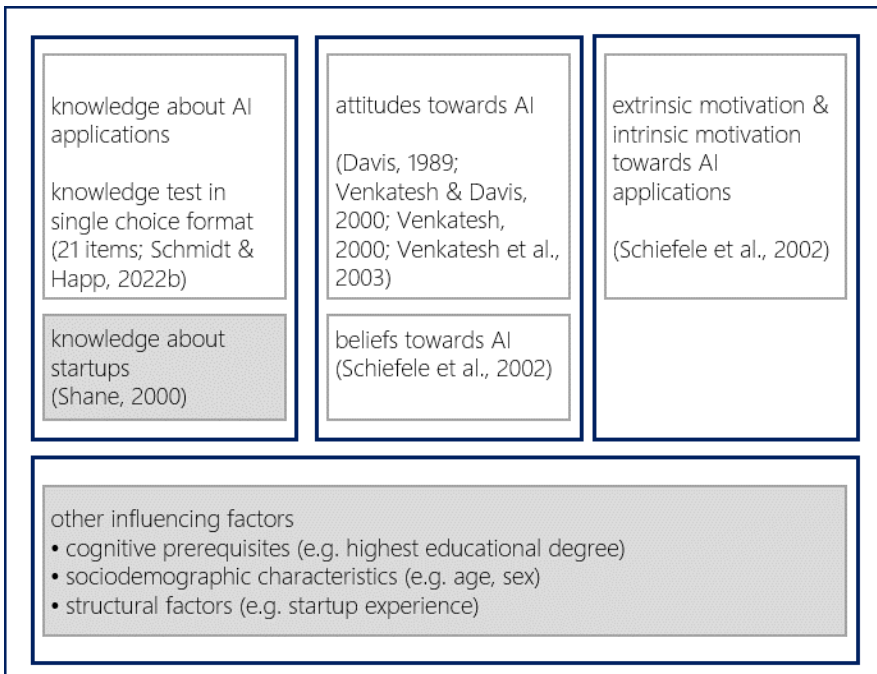


Figure 1: Framework for Facets of AI-Related Entrepreneurial Competence (Adapted from Schmidt & Happ, 2022b)¹

¹ The term AI literacy often is used interchangeably with AI competence. In a comprehensive literature review, Laupichler, Aster, Schirch, and Raupach (2022) found that although the term AI literacy is defined in different ways, one of the most mentioned facets of AI literacy is AI knowledge.

2.2.1 Content knowledge

Research on vocational competence in general has provided ample evidence that cognitive factors such as content knowledge are powerful predictors of achievement (Nitzschke, Velten, Dietzen & Nickolaus, 2019). Therefore, knowledge about AI and AI applications is a central prerequisite for successful AI startups. Such knowledge includes understanding of basic functionalities of AI technologies and AI-based applications (Schmidt & Happ, 2022b).

2.2.2 Attitude toward AI

In a survey on attitudes toward AI conducted in 2021 by the Statista Research Department (2022), 51 % of respondents reported having a positive attitude toward AI. In 2019, 46 % of respondents answered similarly. Schmidt and Happ (2022b) point out that attitude toward AI seldom has been addressed in scientific studies and that instruments for valid measurement have not been developed.

2.2.3 Beliefs about AI

The term “beliefs” goes hand in hand with terms such as values, attitudes, and opinions, a clear definition is missing. According to Pajares (1992) and Richardson (1996), beliefs are enduring ideas, hypotheses, subjective opinions, and assumptions about something (Dubberke, Kunter, McElvany, Brunner & Baumert, 2008). Beliefs are seen as (unconscious) drivers of behavior (Pajares, 1992). There is evidence that beliefs influence judgements on learning (Mueller & Dunlosky, 2017); therefore, beliefs about AI are expected to influence learning about AI.

2.2.4 Motivation related to AI and training motivation

Motivation is relevant in the present study in two respects. First, we consider motivational dispositions in relation to AI (intrinsic and extrinsic motivation), second, we include training motivation. Intrinsic motivation can be expected to influence learning processes positively and thus lead to sustainable learning success and ultimately support implementation (intention to found an AI startup, actual founding) of the content learned (AI and startup knowledge) (Schiefele, 1993, 1996; Schiefele & Köller, 2001). Extrinsic motivation is more about external reasons for engaging with AI applications (e. g. to earn a lot of money); it is not expected to influence the learning processes in the program.

2.3 Criteria for successful training and levels of training evaluation

Criteria for designing and implementing effective training programs have been identified as follows and were taken into account (Salas & Cannon-Bowers, 2001; Hochholdinger, Rowold & Schaper, 2008):

- (1) The relevance of the topic was emphasized in the training offer, which is expected to have a positive effect on the training motivation (Gräsel, Fussangel & Schellenbach-Zell, 2008; Lipowsky, 2011; Lipowsky & Rzejak, 2021).
- (2) The state of research on effective trainings was taken into account, especially input and application phases were offered alternately (Cademartori et al., 2017; Krille, 2019). In addition, reflection was repeatedly encouraged (Cademartori et al., 2017; Lipowsky, 2009).
- (3) Previous findings with regard to the duration of trainings do not yet allow any clear conclusions to be drawn (ibid.), but it is to be expected that very short measures are unlikely to be effective in the long term (Lipowsky & Rzejak 2021; Seifried & Wuttke, 2017). Therefore, the training was designed over a period of several weeks.

The effectiveness of training programs should be evaluated at different levels (Kirkpatrick, 1998; Kirkpatrick & Kirkpatrick, 2006): participants' reactions to the program (e. g., did the participants enjoy the program and find it valuable?), learning gains (e. g., did the participants increase their knowledge?), the behavior (e. g., did the participants use their acquired knowledge, e. g., to write a business plan), as well as the results (e. g., did the participants found a startup and, if so, with what success?)

In this paper we focus on the reaction and the learning success (knowledge about AI applications, knowledge about startups, motivation toward AI, attitudes toward AI and beliefs about AI).

3 The present study

To address the need for an ecosystem for AI talents and for effective training for potential entrepreneurs, we established the H-Ventures training program. This program offers first-time entrepreneurs the opportunity to gain essential knowledge about AI and starting a business, to build valuable networks through mentoring from business experts and working with peers, to build a viable team, and to build a foundation for AI- or data-driven businesses (TechQuartier, 2022). The training program is funded by the state of Hesse,² Germany (from May 2022 to May 2024) and offered through an experienced training provider located in Hesse who advertised the program across its networks and through accelerators at universities. The program, its objectives, learning formats and exact structure will be described in detail in another article (Hangen, Wuttke & Weber, 2024).

² "Digitales Hessen" funding in accordance with the guidelines of the state of Hesse on innovation funding part II, Nr. 1 from December 9, 2016 (StAnz. 52/2016, S. 1676)

The nine-week program is divided into three phases:

- (1) team formation
- (2) bootcamp
- (3) incubation

During the team formation phase participants are encouraged to validate their business idea and build a team. During the bootcamp phase participants become acquainted with other potential founders and mentors. During the incubation phase participants are mentored and coached by experts in their respective fields. In terms of instruction, the program is delivered in a variety of (learning) formats that can be classified as social events (e. g., networking opportunities) and educational sessions (e. g., workshops and input sessions). During social events, participants can exchange ideas and engage in networking; during educational sessions, various issues are addressed such as effective teamwork, agile frameworks, ethics in AI, data product development, business modelling and positioning, and fundraising. Most of the educational sessions are offered in a workshop format and lectures are complemented with opportunities for direct exchange among participants. To accommodate differences in terms of participants' level of knowledge about fundamentals of AI and skills in programming, part of the training program is also offered online and permits participants to work independently at their own pace and at a time that is convenient to them.

Although applicants do not need to have an idea for a new business, they must go through an application process to enroll in the program that involves attending an interview and writing a brief report on their reason for applying to participate in the program. There were no formal restrictions regarding enrolment. Enrolment is free, but participants are expected to commit themselves to the program, which means investing a significant amount of time and energy participating in sessions, contributing to discussions, and doing evaluations.

Our research questions are as follows:³

- (1) What knowledge do participants have about AI applications upon starting the training program and what knowledge do they gain through the program?
- (2) How do the participants rate their knowledge about starting a new business prior to and after the program?
- (3) What are the applicants' initial beliefs about and attitudes toward AI?
- (4) How do intrinsic and extrinsic motivation for AI applications change over the course of the program?

³ The first four questions are answered with the help of a questionnaire, the last two are answered from interviews.

- (5) Why did the applicants decide to enroll in the training program and what were their expectations?
- (6) Are the participants interested in AI in general and do they plan to integrate AI into their startups?

4 Methods

4.1 Measures and data analyses

By addressing the aforementioned research questions, we investigated whether our training program was able to promote the entrepreneurial competences of potential founders in AI or data-driven businesses. We collected data in two ways: with a survey involving administration of a questionnaire before participation in the program and upon completion of the program, and through semi-structured interviews. The survey procedure is illustrated in Figure 2.

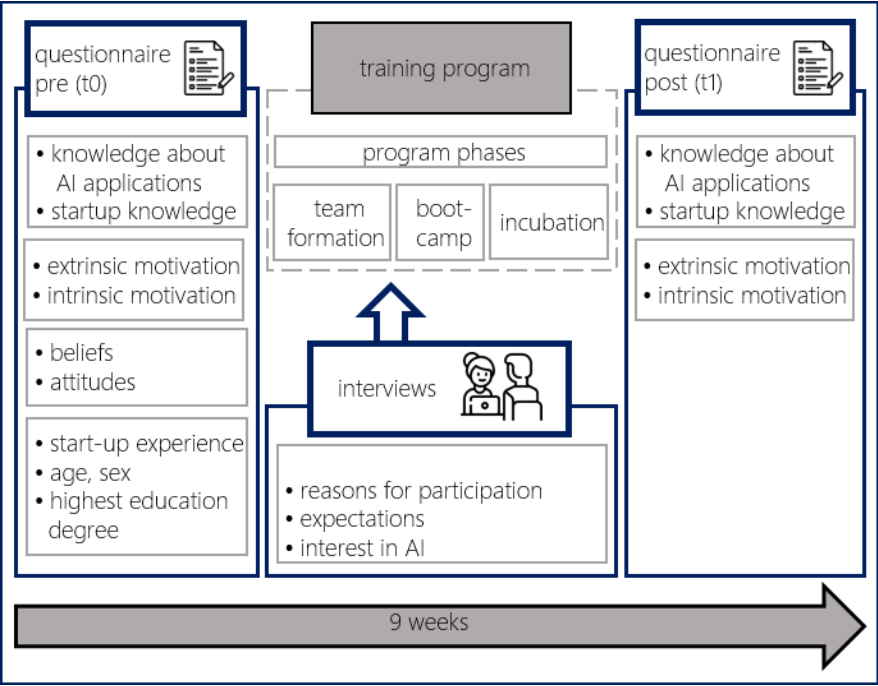


Figure 2: Survey Outline

4.1.1 Questionnaire

The questionnaire was administered using the Unipark survey software program before applicants started the program (pre) and upon completing the program (post). The questions were designed to gather (1) demographic information as well as details regarding their (2) knowledge about AI applications (Schmidt & Happ, 2022a; Schmidt & Happ, 2022b), which were presented on a 21-item knowledge test (single choice format). The assessed knowledge is declarative knowledge in the form of structural or factual knowledge and is differentiated into declarative-static and declarative-procedural knowledge (for further details, see Schmidt in preparation). Motivational scales adapted to the context of AI by Schmidt and Happ (2022b) were used to gather information on participants' (3) intrinsic and extrinsic motivation (Schiefele et al., 2002) (e. g., "I am interested in artificial intelligence because I enjoy working with content related to artificial intelligence."), (4) beliefs about AI (ibid.) (e. g., "There are many scientific findings in the field of artificial intelligence that will always be valid"), and (5) attitudes toward AI (perceived usefulness, perceived ease of use, job relevance) (Davis, 1989; Venkatesh & Davis, 2000) (e. g., "I find it easy to learn how to use artificial intelligence applications.").

To assess participants' self-reported knowledge about starting a business (6), they indicated on a 4-point Likert scale (1 = reject the statement/nothing; 4 = fully agree with the statement/almost everything) how much they knew about various aspects of starting a business (e. g., "what to consider when starting a business in general", Shane, 2000; see also Heinrichs, 2016). Analyses of the responses were conducted using IBM SPSS software.

4.1.2 Interviews

Semi-structured interviews (i. e., warm-up phase, main part, conclusion) were conducted online via Zoom between the first (team formation) and second part (bootcamp) of the program by trained interviewers who followed procedural protocols. The interviews were recorded and transcribed verbatim. The answers to the three interview questions below will be the object of this analysis:

- ▶ Why did you apply to participate in the program?
- ▶ What are your expectations of the program?
- ▶ What is your/your team's business idea?

Following the principles of qualitative content analysis (e. g., Schreier, 2014a) we devised a guide for categorizing responses. The main categories aligned with the research questions. Subcategories were then created based on six interviews as a result of a generalization process (inductive category formation; Mayring, 2010). The interview transcripts were segmented, that is, divided into units for coding prior to initial coding (Schreier, 2014b). To be able to predict the inter coder reliability nine

interviews were coded by two coders ($\kappa = .77$). Categories, subcategories, and the number of codes across all documents are listed in Table 1 (multiple codes per interview were possible).

Table 1: Category system

			Subcategory	Code quantity
1. Initial motivation and expectations	1.1	Reasons for applying to the program	1.1.1 timing/optimal time for respective startup phase	6
			1.1.2 location and network area	14
			1.1.3 recommended by co-founders	5
			1.1.4 searching for new co-founders/team members	19
			1.1.5 finds AI a promising topic/interested in AI	10
			1.1.6 developing new fields or areas	7
	1.2	Expectations of the program	1.2.1 acquisition of knowledge	
			1.2.1.1 knowledge about successful teams	2
			1.2.1.2 unspecified knowledge	15
			1.2.1.3 entrepreneurial knowledge (incl. financing)	17
			1.2.2 networking and visibility within the community	17
			1.2.3 continue working on the idea (motivational boost)	27
			1.2.4 obtain feedback on the startup idea	27
			1.2.5 exchange with other participants (e. g., same problems)	12
2. Startup idea	2.1	Startup sector	2.1.1 finance/fintech	3
			2.1.2 leisure	4
			2.1.3 health & pharma	5
			2.1.4 human resources	2
			2.1.5 sustainability & energy	5
			2.1.6 education	3
			2.1.7 marketing & marketplace	4
			2.1.8 other	4
	2.2	AI	2.2.1 no AI reference	6
			2.2.2 AI reference	26

Notes: AI = Artificial Intelligence

4.2 Sample

We collected data from two cohorts, which we treated as one sample because they had completed the same program only at different times. The questionnaire sample

consisted of applicants to the program; the interview sample consisted of only those applicants who had an idea for their startup when they applied to the program.

Of the 83 respondents to the questionnaire at the first measurement point ($M_{age} = 28.94$, 52 identified as male) 27 could be merged to complete pre/post observations. Almost half of the sample reported having a master's degree, one quarter claimed to have a bachelor's degree, and eight a Ph. D., and 20 stated they had already started a business.

A total of 32 people ($M_{age} = 31.00$, 23 identified as male) were recruited for the interviews. Of these 32 people, seven had a bachelor's degree, 18 a master's degree, and two a Ph. D., and eight indicated that they had already founded a company. The first round of data collection ran from September to November 2022; the second ran from March to May 2023.

5 Results

5.1 Questionnaire

Knowledge of AI applications prior to participation in the program averaged 11.45 points (max = 17; min = 6). Upon completion of the program, this dropped to 10.17 points (max = 17; min = 2). Looking only at those who participated at both measurement points,⁴ a similar trend emerged ($N = 27$, $M_{t0} = 11.26$, $SD = 2.71$; $M_{t1} = 10.33$, $SD = 3.20$), but the difference between the two measurement points was statistically insignificant.

⁴ In Table 2 the mean values of the total sample are shown. For the pre-post comparisons, however, only a subsample was considered, namely the people who participated in both questionnaires.

Table 2: Questionnaire Results

		Mean (standard deviation)	
Facets	Subscales	t0 (N = 84)	t1 (N = 28)
Concerning AI	knowledge	11.45 (2.73)	10.17 (3.51)
	usefulness	3.37 (0.63)	
	attitudes*	2.73 (0.63)	-
	ease of use	3.39 (0.67)	
	job relevance	2.40 (0.44)	-
	beliefs*	2.40 (0.44)	-
	intrinsic	3.21 (0.63)	2.86 (0.55)
	extrinsic	2.97 (0.84)	2.84 (0.83)
	startup knowledge*	2.61 (0.57)	2.81 (0.45)

Notes: *4-point Likert scale; N = sample size; t0 = questionnaire pre; t1 = questionnaire post

In contrast, self-reported knowledge about how to start a business increased from $M = 2.61$ ($SD = 0.57$) to $M = 2.81$ ($SD = 0.45$) when rated on a 4-point Likert scale. Looking at only those who participated at both measurement points, a similar trend emerged ($N = 27$, $M_{t0} = 2.56$ ($SD = 0.51$); $M_{t1} = 2.81$ ($SD = 0.45$)). Self-reported knowledge about how to start a business was significantly higher upon completion of the program than when starting it ($t(26) = 2.32$; $p = .028$; $d = .446$).

Likewise, before the program, there was a significant difference in self-reported knowledge about how to start a business between people who had already founded a company and those who had not ($M_{\text{experienced founder}} = 2.88$ ($SD = 0.49$); $M_{\text{unexperienced founder}} = 2.53$ ($SD = 0.57$); $t(80) = 2.40$; $p = .018$; $d = .62$).

The mean values of attitudes toward AI ranged from $M = 2.73$ ($SD = 0.63$) for ease of use to $M = 3.39$ ($SD = 0.67$) for job relevance and could be considered rather high based on the scale of 1 to 4 used. The mean values of beliefs about AI were $M = 2.40$ ($SD = 0.44$).

In terms of motivation toward AI applications, extrinsic motivation decreased slightly (but not significantly), while intrinsic motivation decreased significantly ($t(23) = 2.63$; $p = .015$; $d = .54$). Intrinsic motivation prior to the program did not differ significantly between people who had already founded a business and those who had not ($M_{\text{experienced founder}} = 3.18$ ($SD = 0.69$); $M_{\text{unexperienced founder}} = 3.20$ ($SD = 0.61$)). Due to the low number of participants after the program ($n = 4$ participants who had already founded a company), the difference could not be calculated, but descriptive differences became apparent ($M_{\text{experienced founder}} = 2.50$ ($SD = 0.29$); $M_{\text{unexperienced founder}} = 2.90$ ($SD = 0.53$)).

5.2 Interviews

From the 32 interviews we conducted, we found 61 utterances regarding why participants chose to enroll in the program. These 61 utterances were grouped into six different reasons for applying to participate in the program. The most common reason was to find co-founders or team members (N = 19):

"That's why I applied to the program, to find co-founders or like-minded people who would like to somehow advance my idea."
(interview 17)

Five people said they had applied because their co-founders had convinced them:

"Because a friend and partner asked me - not pushed me - but asked me if I would like to do this together with him." (interview 8)

Another reason for applying was the fit of the program with the respective startup phase (N = 6, "Then we realized how well the program just fits our next steps."; interview 20) and the expectation that this would give them the best support or encouragement at their current stage of business development.

A frequently cited reason for applying to participate in the program was "location and network area" (N = 14), meaning the program provided a strong startup ecosystem in which participants hoped to build valuable networks. This response was accompanied by interest in AI (AI as promising topic/interest in AI; N = 10), and AI was referred to as "promising topic" and "a field that will grow". Closely related to this response was "developing new fields or areas" (N = 7):

"What we expect from this is to see a little bit what makes other people think, what other ideas there are at the moment. Especially here in the region. Yes, what people want to come up with and implement, where the trends are going." (interview 6)

Reasons for participating in the program were closely aligned with participants' expectations of the program. We found 117 comments relating to expectations and divided them into six categories. The category with the most codings was acquisition of knowledge (N = 34), which was further divided into three subcategories: knowledge about successful teams (N = 2), unspecified knowledge (N = 15), and entrepreneurial knowledge (N = 17). Comments categorized as unspecified knowledge included all those in which participants stated they wanted to build knowledge but did not explain precisely what kind of knowledge. This included statements such as the following:

"Simply first of all, because we always want to take a lot of knowledge with us." (interview 20)

"I went in at the beginning with the expectation that I would just learn a lot." (interview 31)

The utterances concerning knowledge about entrepreneurship often indicated a strong interest in gaining business-related knowledge:

"I want to gain the necessary knowledge, or at least basic knowledge, in order to approach a startup with a certain degree of certainty, so, yes." (interview 13)

"So I'm hoping to get input that will help me when it actually comes to starting the business, in terms of business model and idea development." (interview 14)

In addition to acquiring knowledge, the other main expectation was to receive feedback on one's own startup ideas (N = 27). This may have come from new team members, but the participants mentioned predominantly feedback from mentors and experts in the field.

"My expectations or our expectations would be first and foremost that we get good feedback, especially qualified feedback." (interview 22)

"I would then like to present my idea and get initial feedback. I really want to have this expert opinion to then learn what else do I need to consider? What else do I need to do? Where should I start my company?" (interview 21)

In addition to obtaining feedback, another expectation of the program was to exchange information with other participants (N = 12). Utterances described obtaining social support, exchanging information about similar problems, and finding "like-minded people".

The expectation to network and to generate visibility in the community (N = 17) describes the participants' expectations of the program's positive effects on their professional network of founders of AI or data-driven startups.

"Of course I would like to expand my network!" (interview 21)

Many utterances (N = 27) were assigned to the category "continue working on the idea". Here, participants saw the program as an opportunity to work more deeply and intensively on their idea and, if necessary, to enter into it seriously (e. g., by giving up permanent employment). The program was also described as a motivational boost that should lead to investing more time and resources.

"And I think that's always good when you have the pressure to perform in such a program, so to speak, certain deadlines, certain deliveries, because I think that's pretty cool." (interview 16)

"We met briefly and then he said yes, let's just do it and take the whole thing as a catalyst to take a few steps forward more quickly." (interview 15)

Our sixth research question about whether participants were interested in AI was answered by looking at the participants' startup sectors. As can be seen in Table 1, the startup ideas were spread across a wide range of sectors. Only a small minority did not relate their idea to AI or a data-driven business regardless of their sector. Participants' intentions to incorporate AI into business models varied widely. Some had already developed an idea and wanted to incorporate AI in a subsequent step; others had included AI-based or data-driven business models from the beginning.

Even if the implementation was still in the planning stage for many, ideas were accumulating for the use of Blockchain together with AI, AI in forecasting models, and AI in customer service.

6 Discussion

The special issue topic of learning about AI was considered in terms of founders' competences who intend to have a data-driven startup. We conducted a longitudinal study involving a pre-post survey and semi-structured interviews to track learning with regard to AI-related facets of entrepreneurial competence (see Figure 1). Our data indicate that the participants' knowledge about AI applications did not change despite completing the training program. Although content knowledge is considered a powerful predictor of achievement (Nitzschke et al., 2019), it should be kept in mind that startups are usually founded in teams, and these teams are usually characterized by their diversity. This means that not every team member necessarily has to have in-depth knowledge of AI applications; a basic understanding may be sufficient for some team members. However, our data show that many of the participants in our training program for potential founders did not have even a basic understanding of AI.

As self-assessed startup knowledge increased, it is reasonable to assume that the participants focused more on acquiring general startup knowledge. It is possible that participants were more interested in exploring general startup issues and therefore neglected learning about AI-specific topics addressed in various learning units. Perhaps acquiring startup knowledge was seen as more "hands on" and easier to implement immediately while acquiring AI knowledge was seen more as an on-going endeavor that could be developed at a later time. In addition, the AI knowledge test focusses on the basics of AI applications and therefore may not capture the specific knowledge that participants have acquired. Depending on the focus of the startup, AI knowledge acquisition may have been much more specific, for example, limited to a particular AI model. Although information is available on which learning units were offered, we did not investigate whether or how the content of the units were actually used during or after the training phase.

When looking at the participants' attitudes toward AI, it is important to note that the participants rated the benefits of AI and job relevance as greater than the ease of its use. Thus, they attributed great significance to AI in relation to their job and expressed awareness of the importance of AI. Motivation was divided into extrinsic and intrinsic, and intrinsic motivation was significantly lower after the program, while extrinsic motivation did not decrease significantly. One explanation for the drop in intrinsic motivation could be that participants were confronted with the (technical) difficulties in using AI or they could not implement technological ideas

as they had previously imagined, and so the (expected) fun of working with AI could have suffered as a result.

Analysis of data collected during the interviews reveals that the reasons for enrolling in the training program were manifold. At the top of the list was searching for co-founders or other team members. Further, participants were interested in starting a new business and learning more about AI. Responses to questions regarding expectations of the training program indicate there may be demand for the development of future programs: many saw participation in the program mainly as an opportunity to boost their motivation and their chances for entrepreneurial success. Furthermore, guidance, feedback, and support from experts was another important reason for enrolling in the program, as was the opportunity to build networks. Concerning entrepreneurial ideas, the participants emphasized that they wanted to establish AI in their startups. This finding is consistent with positive attitudes toward AI, beliefs about AI, and motivation toward AI applications.

There were limitations to our study that should be mentioned. First, the sample that completed the pre- and post-questionnaire was small and therefore results cannot be generalized. Second, the sample was drawn from one training program and therefore results must be interpreted within the regional context and within the context of the evaluated program only. Nonetheless, the data offer insight into the AI-related competences of aspiring entrepreneurs and reveal their expectations and motivations. In future training programs for potential founders of AI or data-driven startups, more emphasis should be placed on developing AI competences. While the program seems to have been effective for developing general startup knowledge, the findings of this research and other studies on training should be integrated even more strongly in the design of future programs to make the focus on AI clearer. To this end, (1) the relevance of the topic could be made clearer to increase motivation and (2) the number of implementation phases could be increased, even if they are associated with more effort and increased budget requirements.

In this paper we focused on participants' reactions to their entrepreneurial training program and their learning gains; in future research we will explore the participants' post-training behavior (e. g., whether they apply their newly acquired knowledge) and the results in terms of the impact of their training on their entrepreneurial endeavors (e. g., whether they succeed in setting up a profitable venture). Moreover, we will explore entrepreneurial team building, in particular, the processes of finding appropriate team members, team dynamics, and teamwork.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50 (2), 179-211.
- Alberti, F. (1999). Entrepreneurship education: scope and theory. In C. Salvate, P. Davidsson & A. Persson (Eds.), *Entrepreneurial knowledge and learning: Conceptual advances and directions for further research* (pp. 64-84). Jonkoping: Jonkoping International Business School.
- Bohner, G. (2002). Einstellungen. In W. Stroebe, K. Jonas, M. Hewstone (Hrsg.), *Sozialpsychologie*. Springer-Lehrbuch. Berlin: Springer.
- Brandt, M. (2019). Künstliche Intelligenz rechnet sich [Artificial intelligence pays off]. Verfügbar unter: <https://tinyurl.com/w7z4wk5> [26.07.2023].
- Bundesregierung (2020). Künstliche Intelligenz "Die entscheidende Zukunftstechnologie des 21. Jahrhunderts" [Artificial intelligence "The decisive future technology of the 21st century"]. Verfügbar unter: <https://tinyurl.com/yk3o82> [26.07.2023].
- Cademartori, I., Seifried, J., Wuttke, E., Krille, C., & Salge, B. (2017). Developing a training programme to promote professional error competence in accounting. In E. Wuttke & J. Seifried, *Professional Error Competence of Preservice Teachers: Evaluation and Support* (S. 29-46). SpringerBriefs in Education. Springer, Cham.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13 (3), 319-340.
- Dubberke, T., Kunter, M., McElvany, N., Brunner, M. & Baumert, J. (2008). Lerntheoretische Überzeugungen von Mathematiklehrkräften: Einflüsse auf die Unterrichtsgestaltung und den Lernerfolg von Schülerinnen und Schülern [Learning theory beliefs of maths teachers: influences on lesson planning and students' learning success]. *Zeitschrift für pädagogische Psychologie*, 22 (34), 193-206.
- Ecker, W., Coulon, C-H, Kohler, M. (2021). KI in die Anwendung bringen – Eine Gemeinschaftsaufgabe für Hochschulen, Forschungseinrichtungen, Unternehmen und Politik [Bringing AI into application – a joint task for universities, research institutions, companies and politics] (Whitepaper). München: Plattform Lernende Systeme. Verfügbar unter: <https://www.acatech.de/publikation/ki-in-die-anwendung-bringen/> [07.03.2024]
- European Commission (2006). Entrepreneurship education in Europe: fostering entrepreneurial mindsets through education and learning. Final Proceedings of the Conference on Entrepreneurship Education in Oslo.
- Gaddam, S. (2007). A conceptual analysis of factors influencing entrepreneurship behavior and actions. *ICFAI Journal of Management Research*, 6 (11), 46-63.
- Hangen, Wuttke & Weber (2024). Entwicklung und Evaluation eines Trainingsprogramms für Startups im Bereich data driven ventures [Development and evaluation of a training program for start-ups in the field of data driven ventures]. In preparation.
- Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen. Kompetenzfassung in pädagogischen Handlungsfeldern [Competences as results of educational processes. Assessment of competences in pedagogical fields of action]. *Theorien, Konzepte und Methoden*, 26, 15-26.
- Heinrichs, K. (2016). Dealing with critical incidents in the postformation phase: Design and evaluation of an entrepreneurship education course. *Vocations and Learning*, 9 (3), 257-273.
- Heinrichs, K. (2021). Assessing critical incidents in the post-formation phase: a matter of entrepreneurial expertise? *Journal of International Education in Business*, 14 (2), 161-178.
- Hochholderinger, S., Rowold, J. & Schaper, N. (2008). Ansätze zur Trainings- und Transferevaluation. [Approaches to training and transfer evaluation.] *Trainingsevaluation und Transfersicherung: Modelle, Methoden, Befunde*, 30-52.
- Hoffmann, T. (1999). The meanings of competency. *Journal of European Industrial Training*, 23, 275-286.
- Kirkpatrick, D. (1998). Evaluating training programs: The four levels. San Francisco, CA: Berrett-Koehler.
- Kirkpatrick, D., & Kirkpatrick, J. (2006). Evaluating training programs: The four levels. Berrett-Koehler Publishers.

- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). Zur Entwicklung nationaler Bildungsstandards [Development of national educational standards] (Expertise). Bonn: Bundesministerium für Bildung und Forschung (BMBF).
- Kreditanstalt für Wiederaufbau (KfW) (2015). KfW Gründungsmonitor [KfW Start-up Monitor]. Frankfurt am Main: Kreditanstalt für Wiederaufbau. Verfügbar unter: <https://tinyurl.com/yw7dgz47> [26.07.2023].
- Kuratko, D. F. (2005). The emergence of entrepreneurship education: development, trends, and challenges. *Entrepreneurship Theory and Practice*, 29 (5), 577-598.
- Laupichler, M. C., Aster, A., Schirch, J. & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101.
- Maresch, D., Harms, R., Kailer, N. & Wimmer-Wurm, B. (2016). The impact of entrepreneurship education on the entrepreneurial intention of students in science and engineering versus business studies university programs. *Technological Forecasting & Social Change*, 104, 172-179.
- Mayring, P. (2010). Qualitative Inhaltsanalyse [Qualitative content analysis]. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Moberg, K. (2014). Two approaches to entrepreneurship education: The different effects of education for and through entrepreneurship at the lower secondary level. *The International Journal of Management Education*, 12 (3), 512-528.
- Mueller, M. L. & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, 93, 245-258.
- Nabi, G., Holden, R. & Walmsley, A. (2006). Graduate career-making and business start-up: a literature review. *Education+Training*, 48 (5), 373-385.
- Nitzschke, A., Velten, S., Dietzen, A. & Nickolaus, R. (2019). Relevance of professional experience and vocational knowledge for solving problems – an analysis of perspective electrical engineering technicians. *Zeitschrift für Weiterbildungsforschung*, 43, 117-131.
- Norris, N. (1991). The trouble with competence. *Cambridge Journal of Education*, 21, 331-341.
- Oosterbeek, H., van Praag, M. & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54, 442-454.
- Pajares, M. (1992). Teachers' beliefs and educational research. Cleaning up a messy construct. *Review of Educational Research*, 62 (3), 307-332.
- Plümer, T. & Niemann, M. (2016). Existenzgründung Schritt für Schritt. Mit 2 ausführlichen Businessplänen (2. aktual. Aufl.) [Business start-up step by step. With 2 detailed business plans]. Wiesbaden: Gabler.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. *Handbook of research on teacher education*, 2 (102-119), 273-290.
- Roxas, B. G., Cayoca-Panizales, R., Mae de Jesus, R. (2008). Entrepreneurial knowledge and its effects on entrepreneurial intentions: Development of a conceptual framework. *Asia-Pacific Social Science Review*, 8 (2), 61-77.
- Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual review of psychology*, 52 (1), 471-499.
- Schiefele, H. (1993). Brauchen wir eine Motivationspädagogik? [Do we need a motivational pedagogy?]. *Zeitschrift für Pädagogik*, 39 (2), 177-186.
- Schiefele, U. (1996). Motivation und Lernen mit Texten [Motivation and learning with texts]. Göttingen: Hogrefe.
- Schiefele, U. & Köller, O. (2001). Intrinsische und extrinsische Motivation [Intrinsic and extrinsic motivation]. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 304-310). Weinheim: Beltz.
- Schiefele, U., Moschner, B. & Husstegge, R. (2002). Skalenhandbuch SMILE-Projekt [Scales manual SMILE project]. Bielefeld: Universität Bielefeld, Abteilung für Psychologie.

- Schmidt, J. & Happ, R. (2022a). Befunde aus einem Lehr- und Forschungsprojekt zur Integration von Inhalten zu KI-Anwendungen bei angehenden Berufsschullehrkräften [Findings from a teaching and research project on the integration of learning on AI applications for prospective vocational school teachers]. In D.-K. Mah & C. Torner (Hrsg.), *Künstliche Intelligenz mit offenen Lernangeboten an Hochschulen lehren. Erfahrungen und Erkenntnisse aus dem Fellowship-Programm des KI-Campus* (S. 57-72). Berlin: KI-Campus.
- Schmidt, J. & Happ, R. (2022b). Fachspezifisches Wissen, Einstellungen und Überzeugungen zu Anwendungen der Künstlichen Intelligenz bei Studierenden der Wirtschaftspädagogik [Subject-specific knowledge, attitudes and beliefs about applications of artificial intelligence among students of business education], *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung*, 2022, 89-102.
- Schreier, M. (2014a). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. [Variants of qualitative content analysis: A guide through the jungle of terminology.] *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 15 (1).
- Schreier, M. (2014b). Qualitative content analysis. In U. Flick (Ed.), *The SAGE Handbook of qualitative data analysis*. London: SAGE.
- Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organ Sci*, 11 (4), 448-469.
- Short, E. C. (1985). The concept of competence: Its use and misuse in education. *Journal of Teacher Education*, 36, 2-6.
- Siegfried, C., & Wuttke, E. (2021). Development of economic competence of students at general grammar schools through a non-formal opportunity to learn with experts. *Citizenship, Social and Economics Education*, 20 (2), 122-142.
- Stamboulis, Y. & Barlas, A. (2014). Entrepreneurship education impact on student attitudes. *The International Journal of Management Education*, 12, 365-373.
- Statista Research Department (2022). Umfrage zur Einstellung gegenüber Künstlicher Intelligenz 2021 [Survey on attitudes towards artificial intelligence 2021]. Verfügbar unter: <https://tinyurl.com/ys4grohd> [26.07.2023].
- Seenivasan, A. & Suresh, M. (2023). Twenty years of entrepreneurship education: a bibliometric analysis. *Entrepreneurship Education*, 6 (1), 45-68.
- TechQuartier (2022). Find your co-founder and launch your data-driven venture. Retrieved from: <https://h-ventures.studio/> [26.07.2023].
- Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46 (2), 186-204.
- Wang, X., Yang, H., Han, H., Huang, Y. & Wu, X. (2022). Explore the entrepreneurial process of AI start-ups from the perspective of opportunity. *Systems Research on Artificial Intelligence*, May/June 2022, 569-580.
- Wu, W. Y., Chang, M. L. & Chen, C. W. (2008). Promoting innovation through the accumulation of intellectual capital, social capital, and entrepreneurial orientation. *R&D Management*, 38 (3), 265-277.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit [Comparative performance assessments in schools – a controversial matter of course]. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-32). Weinheim: Beltz.
- Widding, L. (2005). Building entrepreneurial knowledge reservoirs. *Journal of Small Business and Enterprise Development*, 12 (4), 595-615.
- Zlatkin-Troitschanskaia, O. & Seidel, J. (2011). Kompetenz und ihre Erfassung. Das neue „Theorie-Empirie-Problem“ der empirischen Bildungsforschung? [Competence and its assessment. The new "theory-empiricism problem" of empirical educational research?] In O. Zlatkin-Troitschanskaia (Hrsg.), *Stationen empirischer Bildungsforschung. Traditionslinien und Perspektiven* [Klaus Beck zum 70. Geburtstag gewidmet] (S. 218-233). Wiesbaden: VS, Verlag für Sozialwissenschaften.

Autorinnen

Jule Hangen, Wirtschaftspädagogik, insb. empirische Lehr-Lern-Forschung, Goethe Universität Frankfurt

Prof. Dr. Eveline Wuttke, Wirtschaftspädagogik, insb. empirische Lehr-Lern-Forschung, Goethe Universität Frankfurt

Korrespondenz an: hangen@econ.uni-frankfurt.de

Sabine Seufert, Lukas Spirgi, Jan Delcker, Joana Heil & Dirk Ifenthaler

Umgang mit KI-Robotern: maschinelle Übersetzer, Textgeneratoren, Chatbots & Co – Eine empirische Studie bei Erstsemester-Studierenden

Künstliche Intelligenz (KI) wirkt sich zunehmend auf das Leben der Menschen aus, indem sie zu einer verstärkten Interaktion zwischen Menschen und Maschine führt (Kim, 2022). KI-Roboter werden als Agenten betrachtet, die durch KI-Programmierung Aufgaben übernehmen, die traditionell von Menschen ausgeführt wurden. Es gibt zwei Kategorien von KI-Robotern: solche mit menschenähnlicher Leistung und solche mit menschenähnlicher Erscheinung zur sozialen Interaktion. Basierend auf diesem Konzept wurden für diese Studie sieben konkrete Anwendungsfälle entwickelt. Eine Online-Umfrage unter Studierenden (N = 636) im September 2022 (kurz vor der Veröffentlichung von ChatGPT) untersuchte sowohl die Nutzungshäufigkeit als auch die ethische Bewertung der Studierenden für jeden der sieben Anwendungsfälle. Zudem wurde sowohl ein Risiko- als auch ein Chancenindex gebildet, welche die generelle Wahrnehmung der Studierenden gegenüber KI erheben. Die wahrgenommenen Risiken übersteigen die erkannten Chancen. Die Nutzungshäufigkeit von KI-Roboter-Typen durch Studierende ist insgesamt als eher gering einzustufen. Die Akzeptanz ethischer Grundsätze unterscheidet sich abhängig vom Typ des KI-Roboters, wobei Studierende insbesondere bei KI-Robotern zur (menschenähnlichen) Textgenerierung eher Wert auf Gerechtigkeit und bei KI-Robotern für menschenähnliche, soziale Interaktionen auf Transparenz legen.

Schlagwörter: KI-Ethik – KI-Kompetenz – KI-Nutzung – KI-Roboter – Künstliche Intelligenz (KI)

Interaction with AI robots: machine translators, text generators, chatbots & co - an empirical study among first-year students

Artificial Intelligence (AI) increasingly influences people's lives by fostering enhanced interaction between humans and machines (Kim, 2022). AI robots are considered agents that undertake tasks traditionally performed by humans through AI programming. There are two categories of AI robots: those with human-like performance and those with human-like appearance for social interaction. Based on this concept, seven specific use cases were developed for this study. An online survey conducted among students (N = 636) in September 2022 (shortly before the release of ChatGPT) examined both the frequency of usage and the ethical evaluation of each of the seven use cases by students. Additionally, a risk and an opportunity index were formed to assess students' general perception of AI. The perceived risks outweigh the recognized opportunities. The frequency of students' usage of AI robot types is generally considered relatively low. The acceptance of ethical principles varies depending on the kind of AI robot, with students particularly valuing fairness for AI robots involved in (human-like) text generation and transparency for AI robots engaged in human-like social interactions.

Keywords: AI Ethics – AI Literacy – AI Robots – AI Usage – Artificial Intelligence (AI)

1 Einleitung

Die Künstliche Intelligenz (KI) scheint zunehmend das Leben der Menschen durch mehr Interaktionen zwischen Menschen und Robotern zu beeinflussen (Kim, 2022). KI bezeichnet Systeme, die auf Verfahren maschinellen Lernens, logischer Programmierung oder statistischer Verfahren basieren und ausgehend von der Analyse von Daten Inhalte, Vorhersagen, Empfehlungen, Entscheidungen hervorbringen oder Handlungen ausführen können (Europäische Kommission, 2022).

Seit der Einführung von ChatGPT im November 2022 hat die KI in Bereichen wie Arbeit und Bildung Einzug gehalten (Garrel, Mayer & Mühlfeld, 2023). Die Leistungsfähigkeit dieser KI hat in der Hochschullehre zu verbreiteten Bedenken geführt, insbesondere hinsichtlich der potenziellen Nutzung durch Studierende zum Plagiierten, indem sie KI-generierte Texte für unkontrolliert erstellte Arbeiten verwenden (Cotton, Cotton & Shipway, 2023). In öffentlichen Diskussionen dominiert daher oft die Perspektive der Lehrenden und der Hochschulverwaltung (Alshami, Elsayed, Ali, Eltoukhy & Zayed, 2023). Vor diesem Hintergrund verfolgt dieser Beitrag das Ziel, den Umgang mit KI in der Hochschullehre aus der Sicht der Studierenden zu erforschen. Dabei ist es zunächst von zentraler Bedeutung, wie Studierende die Möglichkeiten und Gefahren der KI in der Gesellschaft einschätzen, was für einen verantwortungsbewussten Einsatz von KI entscheidend zu sein scheint (Kong, Man-Yin Cheung & Zhang, 2021). Bisher existiert eine begrenzte Anzahl empirischer Studien zur Nutzung von KI in der Hochschullehre (Garrel, et al., 2023; Kim, 2022; Lim, Gunasekara, Pallant, Pallant & Pechenkina, 2023). Darüber hinaus existieren bisher keine konzeptionellen Rahmenwerke, um KI-Roboter zu definieren, die sowohl hinsichtlich ihrer Leistung als auch in ihrer Interaktionsform immer menschenähnlicher werden (Dang & Liu, 2022; Roesler, Manzey & Onnasch, 2021). Durch diesen Ansatz wird es möglich, die Nutzungshäufigkeit sowie die Einstellungen zur ethischen Nutzung von KI anhand spezifischer Anwendungsfälle genauer zu analysieren – ein Aspekt, der bisher noch unerforscht ist und von Forschenden im Bereich der KI-Ethik zunehmend gefordert wird (Jobin, Ienca & Vayena, 2019). An dieser Forschungslücke setzt daher der vorliegende Beitrag an.

Der Beitrag ist folgendermaßen aufgebaut: Zuerst wird in Kapitel 2 der Begriff ‚KI-Roboter‘ definiert und die verschiedenen Typen von KI-Robotern kategorisiert, um die Formen der Menschenähnlichkeit von KI hervorzuheben. Zudem wird dargelegt, nach welchen Kriterien eine ethische Nutzung von KI beurteilt werden kann. Im nachfolgenden Kapitel 3 werden sowohl die Forschungsfragen als auch das Konzept der durchgeführten Studie erörtert. Kapitel 4 befasst sich mit dem methodischen Ansatz der empirischen Untersuchung. Die Präsentation der Untersuchungsergebnisse erfolgt in Kapitel 5. In Kapitel 6 findet eine detaillierte Diskussion der Ergeb-

nisse statt. Den Abschluss bildet Kapitel 7 mit Implikationen für die Hochschullehre und einem Ausblick auf zukünftige Forschungen.

2 Theoretischer Hintergrund

2.1 KI-Roboter und neue Mensch-Maschine Interaktionen

Die Begriffe ‚KI-Technologien‘, ‚KI-Tools‘ und ‚KI-Systeme‘ werden oft synonym verwendet und dienen als Sammelbezeichnungen für Softwareanwendungen, die auf Künstlicher Intelligenz basieren, um spezifische Aufgaben zu automatisieren, zu optimieren oder zu vereinfachen. Im Kontext der KI hat sich zudem der Begriff des Agenten (Akteur) als zentral erwiesen (Murphy, 2019). Ein ‚Intelligenter Agent‘ wird dabei als ein System definiert, das seine Umgebung wahrnimmt und Handlungen durchführt, mit dem Ziel, seine Erfolgswahrscheinlichkeit zu maximieren (Murphy, 2019). Es kann somit auch Veränderungen in der Umgebung bewirken. Die Betrachtung der KI als (teil-)autonome, lernfähige Agenten rückt ethische Aspekte und die Notwendigkeit einer verantwortungsbewussten Nutzung dieser Technologie in den Fokus (Floridi & Cows, 2019).

Ein KI-basierter oder intelligenter Roboter wird ursprünglich in der Robotikforschung als ein physisch situierter Agent in der realen Welt definiert (Murphy, 2019). Vor dem Hintergrund der neuesten Entwicklungen im Bereich des Natural Language Processing (NLP) der KI gewinnen Conversational Agents zunehmend an Bedeutung, die eine Interaktion über Text- oder Sprachschnittstellen mittels natürlicher Sprache ermöglichen und menschliche Gesprächspartner simulieren (Winkler, Söllner & Leimeister, 2021). Conversational Agents sind Chatbots und können in Software-Anwendungen wie Online-Plattformen oder digitale Assistenten integriert oder über Schnittstellen in die Hardware von Robotern eingebunden werden (Wollny et al., 2021). Infolge dieser Fortschritte wird der Begriff des KI-Roboters in neueren Forschungsstudien umfassender definiert: als eine Künstliche Intelligenz, die entweder eine menschenähnliche Erscheinung aufweist oder menschenähnliche Leistungen erbringt (Dang & Liu, 2022; Roesler et al., 2021).

Menschenähnliche Leistungen können hinsichtlich des Umfangs und der Komplexität ihrer Funktionalitäten charakterisiert werden (Dang & Liu, 2022). Im Bereich der KI-basierten Textgenerierung im akademischen Kontext kann das Stufenmodell von Boyd-Graber, Okazaki und Rogers (2023) als Referenzrahmen dienen. Für die Erstellung wissenschaftlicher Arbeiten hat die Association for Computational Linguistics, eine internationale Forschungsgemeinschaft, die sich mit Sprachmodellen wie ChatGPT beschäftigt, Richtlinien für den ethischen Umgang mit KI-basierten Schreibtools herausgegeben (Boyd-Graber et al., 2023). Innerhalb dieses Orientierungsrahmens lassen sich verschiedene Stufen definieren, die einen zunehmenden

Leistungsumfang der KI in der Texterstellung aufzeigen, die den Neuigkeitsgehalt der generierten Inhalte bestimmen:

- 1) Übersetzungsleistung, reine Sprachunterstützung,
- 2) Unterstützung bei der Eingabe von Kurztexten (Schreibassistenz),
- 3) Text mit geringem Neuigkeitsgrad (z. B. Zusammenfassen, paraphrasieren) und
- 4) Text mit neuen Ideen, hoher Neuigkeitswert der Inhalte.

Der Referenzrahmen enthält für jede Stufe Richtlinien, ob der Einsatz von KI bei der Texterstellung deklariert werden muss. Grundsätzlich gilt: Je höher der Leistungsumfang der KI, desto wichtiger ist die Deklaration und die exakte fachliche Prüfung der von der KI erstellten Texte.

Menschenähnliche Erscheinung von KI-Robotern, die neue Dimensionen der Mensch-Maschine-Interaktion eröffnen, lassen sich nach ihrer Modalität charakterisieren. Dies definiert die parallele Nutzung verschiedener Sinneskanäle zur Informationsübermittlung (Minge, 2012). Ein Mediensystem gilt dann als multimodal, wenn es mehr als einen Sinneskanal für die Interaktion verwendet. Solche Systeme fördern natürlichere Interaktionsformen, was die Effizienz der Nutzung, damit den Erfolg der Kommunikation (Weidemann, 2002) und u. U. sogar den Lernerfolg (Wang, Pang, Wallace, Wang & Chen, 2022) steigert. Während Chatbots die textuelle und auditive Modalität für ihre digitale Identität bedienen, repräsentieren Avatare ihre virtuelle Identität, indem sie Gestik, Mimik und Bewegungen nutzen und auch eine reale Person nachbilden können. Im Gegensatz zu Chatbots und Avataren (Software), haben soziale Roboter auch eine physische Einheit (Software und Hardware), sind somit in der physischen Welt verankert und bieten auch eine haptische Dialogschnittstelle (vgl. Abbildung 1). Sie nehmen ihre Umgebung über Sensoren wahr, interagieren mittels Motorik und können auch Emotionen zeigen (Guggemos, Seufert, Sonderegger & Burkhard, 2022).

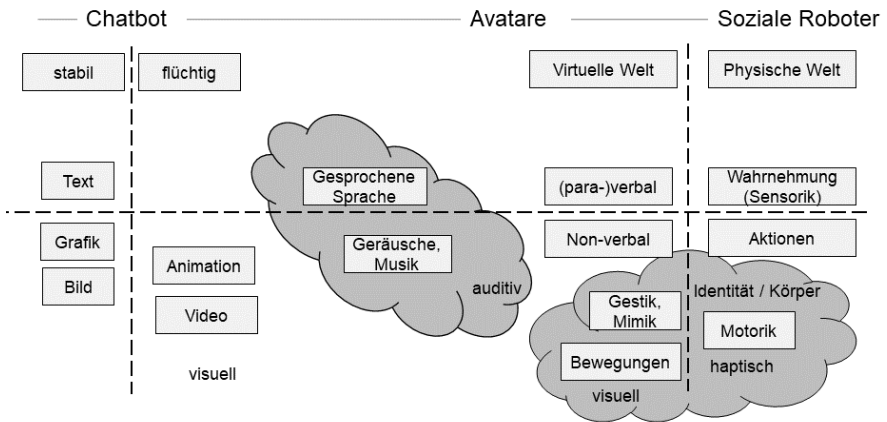


Abbildung 1: Rahmenkonzept von KI-Robotern (eigene Darstellung)

Zusammenfassend lässt sich festhalten: ‚KI-Technologien‘, ‚KI-Tools‘ und ‚KI-Systeme‘ dienen als allgemeine Oberbegriffe. ‚Intelligente Agenten‘ hingegen bezeichnet eine spezifischere Kategorie, die Systeme umfasst, welche ihre Umgebung wahrnehmen und Handlungen mit dem Ziel ausführen, ihre Erfolgswahrscheinlichkeit zu maximieren. Der Begriff ‚KI-Roboter‘ betont die Menschenähnlichkeit. Ursprünglich bezog sich dieser auf das äußere Erscheinungsbild im Sinne einer physischen Präsenz, doch zunehmend wird er auch verwendet, um eine menschenähnliche Leistung und/oder Erscheinungsform zu beschreiben. In der öffentlichen Berichterstattung wird die Software ChatGPT oft als ‚Roboter‘ bezeichnet, wie zum Beispiel im NZZ-Artikel von Kaeser (2022), in dem es heißt: „ChatGPT: Der Roboter schreibt nicht, er schwafelt“.

2.2 Ethischer Umgang mit KI

Die Fortschritte in der Künstlichen Intelligenz bringen sowohl beträchtliche Möglichkeiten als auch ernstzunehmende Herausforderungen mit sich, die insbesondere eine ethisch verantwortungsvolle Anwendung der KI verlangen. Bao et al. (2022) entwickelten einen Index zur Bewertung der potenziellen Vorteile und Gefahren der KI. Zu den Vorteilen gehören unter anderem die Verbesserung der individuellen Gesundheit und die Verringerung von Verzerrungen in menschlichen Entscheidungen. Die Gefahren beinhalten jedoch Aspekte wie die Verstärkung sozialer Ungleichheiten oder eine Zunahme der Arbeitslosigkeit, bedingt durch den Ersatz menschlicher Arbeitskräfte durch Maschinen.

Aufgrund des großen Einflusses von KI auf die Gesellschaft haben viele Organisationen Initiativen gestartet, um ethische Grundsätze für die Anwendung von KI zu etablieren. Diese ethischen Leitlinien sollen eine verantwortungsvolle Gestaltung und Anwendung von KI gewährleisten, damit die Technologie positive Auswirkungen auf die Gesellschaft entfaltet (Floridi & Cowls, 2019). Jobin et al. (2019) haben in ihrer Meta-Studie bestehende Ethikrichtlinien für KI untersucht und miteinander verglichen. Sie erstellten eine Übersicht über aktuelle Prinzipien und Richtlinien für eine ethische KI, um zu prüfen, ob eine weltweite Konvergenz sowohl in den Prinzipien der ethischen KI als auch in den Anforderungen für ihre Implementierung erkennbar ist. Ihre Analyse zeigte, dass sich eine globale Übereinstimmung hinsichtlich fünf ethischer Prinzipien abzeichnet:

- ▶ **Transparenz:** Es gibt Forderungen nach klarer Offenlegung darüber, wie KI-Systeme Entscheidungen treffen. Solche Systeme sollten so konzipiert sein, dass ihre Arbeitsweise und Entscheidungsprozesse nachvollziehbar und transparent sind. Dies fördert das Vertrauen und ermöglicht es den Nutzern und Betroffenen, das System zu verstehen und gegebenenfalls zu hinterfragen. Zudem ist damit verbunden, die Nutzung transparent darzustellen und die menschliche Verantwortlichkeit in der Nutzung klar zu definieren.
- ▶ **Gerechtigkeit und Fairness:** Datenverzerrungen, bei denen bestimmte Muster aus den Trainingsdaten, die zwar mit dem Ergebnis korrelieren, aber zu Verzerrungen oder Voreingenommenheit führen, prägen das Verhalten des Algorithmus. Werden verzerrte Datensätze für das Training von KI-Systemen verwendet, kann das System diese Verzerrungen – beispielsweise Diskriminierung von Minderheiten – in seinen Handlungen widerspiegeln oder sogar verstärken. Die verwendeten Algorithmen beruhen auf Regeln, die eine bewusste oder unbewusste Bevorzugung bestimmter Personen oder Handlungsoptionen zur Folge haben können. Die Verhinderung von Diskriminierung und die Gewährleistung, dass der KI-Einsatz nicht zu sozialen Ungleichheiten führt, sind daher zentral.
- ▶ **Datenschutz und Privatsphäre:** Der Schutz der Privatsphäre und der persönlichen Daten von Individuen ist essenziell. Für das Training von KI-Systemen werden große Mengen an Daten benötigt. Dadurch, dass immer mehr Daten gesammelt und effizienter miteinander verknüpft werden können, entsteht die Möglichkeit, aus scheinbar unpersönlichen Daten Rückschlüsse auf einzelne Personen zu ziehen. Dies führt zu neuen Herausforderungen im Bereich des Datenschutzes und der Wahrung der Privatsphäre.
- ▶ **Nicht-Schädigung und Solidarität:** Die Entwicklung und der Einsatz von KI-Systemen sollten darauf ausgerichtet sein, Schaden zu vermeiden, das allgemeine Wohl sowie den gesellschaftlichen Nutzen zu fördern und sicherzustellen, dass KI-Systeme die menschliche Autonomie nicht beeinträchtigen.

- ▶ **Verantwortliche Entwicklung von KI:** Dieses Prinzip richtet sich an die Gestaltung und Qualitätsentwicklung von KI-Systemen. Es muss definiert werden, wer für die Konsequenzen mangelhafter KI-Systeme verantwortlich ist. Damit kann das Vertrauen in KI-Systeme durch die Entwicklung zuverlässiger und sicherer Technologien gestärkt werden.

Die Ethikrichtlinien und skizzierten Prinzipien sind in einer sehr allgemeinen Form gehalten. Die Forschungsgruppe um Jobin et al. (2019) betont deshalb die Notwendigkeit, diese ethischen Richtlinien speziell für KI-Systeme und ihre Anwendungsgebiete zu konkretisieren. Die ethischen Aspekte variieren je nach Interessengruppe, und es ist wichtig, die jeweilige Perspektive einzunehmen, um sich mit diesen ethischen Fragen auseinanderzusetzen. Beispielsweise könnten Bildungseinrichtungen die Auswirkungen von KI-Technologien auf den Lernprozess und die Gerechtigkeit im Bildungswesen betrachten, während Softwareentwickler und KI-Forschende sich mit Themen wie der Transparenz von Algorithmen und der Vermeidung von Voreingenommenheit in KI-Systemen beschäftigen könnten. Datenschutzbeauftragte würden sich hingegen auf den Schutz der Privatsphäre und den sicheren Umgang mit Nutzerdaten konzentrieren. Jede dieser Gruppen hat eine eigene Perspektive auf ethische Fragen, die es zu berücksichtigen gilt (Morley et al., 2023).

3 Die vorliegende Studie

Diese Studie verfolgt das Ziel, den Umgang mit KI-Robotern in der Hochschulbildung aus der Perspektive der Studierenden zu beleuchten. Insbesondere soll untersucht werden, wie häufig die Studierenden KI-Roboter nutzen und wie sie die Chancen und Risiken von KI-Robotern einschätzen. Die Untersuchung der Wahrnehmung von Studierenden bezüglich der Möglichkeiten und Gefahren, die KI in der Gesellschaft mit sich bringt, ist von großer Wichtigkeit, da sie das Fundament für einen verantwortungsvollen Einsatz von KI bildet (Kong et al., 2021). Es wird davon ausgegangen, dass die Fähigkeit der Studierenden, KI aus verschiedenen Blickwinkeln zu beurteilen, entscheidend für ihren durchdachten Einsatz ist und wesentlich zur Gestaltung einer von Technologie bestimmten Zukunft beiträgt (Kong et al., 2021). Individuen, die der Meinung sind, dass ihr Umgang mit KI bedeutsame Auswirkungen hat, zeigen tendenziell eine größere innere Motivation, die erforderlichen Fähigkeiten zu erlernen (Frymier, Shulman & Houser, 1996). Daher ist es im universitären Kontext wichtig, interdisziplinäre Diskussionen und Reflexionen zu den Auswirkungen von KI zu fördern.

Auf der konkreten Anwendungsebene erscheint es sinnvoll, die Nutzung und ethische Einstellung nach KI-Roboter Typen in menschenähnliche Leistungsfähigkeiten und menschenähnliche Erscheinung zu differenzieren: Diese Unterscheidung ist von Bedeutung, da beide Aspekte unterschiedliche Facetten der Technologieakzeptanz

haben. Die menschenähnliche Leistung von KI bezieht sich auf deren Fähigkeit, Aufgaben auf eine Art und Weise zu bewältigen, die menschlicher Kognition, Problemlösung und Entscheidungsfindung ähnelt. Im Gegensatz dazu beschreibt die menschenähnliche Erscheinungsform die physische oder visuelle Gestaltung der KI, die die Wahrnehmung und emotionale Reaktion der Nutzer beeinflussen kann. Durch das getrennte Erfassen dieser Merkmale kann untersucht werden, inwiefern jede dieser Eigenschaften die Akzeptanz, das Vertrauen und die Bereitschaft zur Interaktion mit KI-Technologien im Hochschulbereich beeinflusst.

Die nachfolgende Abbildung veranschaulicht das Forschungsdesign. Eine Systematisierung der Typen von KI-Robotern wurde nach der menschenähnlichen Leistung sowie menschenähnlichen Erscheinung unterschieden. Für die Studie wurden insgesamt sieben verschiedene Anwendungsfälle entwickelt. Jeder Anwendungsfall beschreibt einen Typ von KI-Robotern, mit denen die Studierenden in der Hochschulbildung in Kontakt kommen können. Die Anwendungsfälle 1 bis 4 sind KI-Roboter mit menschenähnlicher Leistung. Die Anwendungsfälle 5 bis 7 sind KI-Roboter mit menschenähnlicher Erscheinung. Auf Basis der sieben Anwendungsfälle wurde ein Fragebogen entwickelt (siehe Kapitel 4.2), um die drei definierten Forschungsfragen (FF) zu beantworten. Die in der Studie definierten KI-Roboter mit menschenähnlichen Leistungen werden von den Studierenden vor allem zur Textgenerierung verwendet (FF2) und die definierten KI-Roboter mit menschenähnlicher Erscheinung werden primär für die soziale Interaktion verwendet (FF3). Aus der übergeordneten Zielsetzung der Studie ergeben sich somit die folgenden Forschungsfragen:

- ▶ Wie beurteilen Studierende die potenziellen Chancen und Risiken, die Künstliche Intelligenz für die Gesellschaft mit sich bringt (FF1)?
- ▶ In welchem Umfang setzen Studierende KI-Roboter für die Texterstellung ein und wie beurteilen sie in diesem Kontext die relevanten ethischen Aspekte (FF2)?
- ▶ In welchem Umfang setzen Studierende KI-Roboter für soziale Interaktionen ein und wie beurteilen sie in diesem Kontext die relevanten ethischen Aspekte (FF3)?



Abbildung 2: Forschungsdesign (eigene Darstellung)

Anmerkungen: FF = Forschungsfrage

Die in der Studie von Jobin et al. (2019) entwickelten ethischen Prinzipien zur Nutzung von KI wurden in der vorliegenden Studie auf die beiden Typen von KI-Robotern angewendet und spezifisch für die Hochschullehre abgeleitet.

KI-Roboter Typen mit menschenähnlicher Leistung für die Texterstellung:

- ▶ **Transparenz:** Die mit diesen Hilfsmitteln erstellten Passagen sind deutlich als solche gekennzeichnet, die Eigenständigkeitserklärung am Ende einer schriftlichen Arbeit wird angepasst und spezifiziert die Verwendung solcher Hilfsmittel (damit wird das Ziel verfolgt, die Eigenleistung des Menschen im Verhältnis zur KI zu würdigen),
- ▶ **Gerechtigkeit, Fairness und Gleichheit:** freier Zugang für alle Studierende, um soziale Ungleichheiten durch den Einsatz von KI zu vermeiden; zudem wurde eine weitere Option entwickelt, die eine restriktive Nutzung für eine Zielgruppe untersuchen soll: Ausschließlich Studierende mit Sprachproblemen (Legasthenie) dürfen diese KI benutzen, um Benachteiligungen auszugleichen,
- ▶ **Datenschutz und Privatsphäre:** es werden keine sensiblen Daten gesammelt und gespeichert,
- ▶ **Kein Schaden, Solidarität:** Diese werden nur für Aufgaben verwendet, die nicht benotet werden (damit kann die Vermeidung von Schaden und Risiken in der

Nutzung mangelhafter KI adressiert werden). Da zum Zeitpunkt der Datenerhebung leistungsfähige Textgeneratoren wie ChatGPT noch nicht weit verbreitet waren, wurde der Aspekt, dass die KI keinen Menschen ersetzen würde (Solidarität), nicht erfasst.

KI-Roboter Typen mit menschenähnlicher Gestalt für soziale Interaktionen:

- ▶ Transparenz: bei Chatbots: Es wird klar kommuniziert, dass es sich nicht um eine echte Person handelt; bei Avatar: Wasserzeichen weisen auf Avatar hin; bei Sozialen Roboter: kein menschenähnliches Design (damit es sich offensichtlich von einem Menschen unterscheidet),
- ▶ Gerechtigkeit, Fairness und Gleichheit: freier Zugang für alle Studierende, um soziale Ungleichheiten durch den Einsatz von KI zu vermeiden,
- ▶ Datenschutz und Privatsphäre: es werden keine sensiblen Daten gesammelt und gespeichert,
- ▶ Kein Schaden, Solidarität: Die KI würde keinen Menschen ersetzen (durch den Einsatz von KI-Chatbots zur besseren Betreuung der Studierenden würden zum Beispiel keine menschlichen Arbeitsplätze verloren gehen).

In dieser Studie wurde das fünfte ethische Prinzip nach Jobin et al. (2019) jeweils nicht berücksichtigt, weil es sich hauptsächlich auf die Interessengruppe der Entwickler bezieht, die für eine verantwortungsvolle Herstellung von KI-Technologien zuständig sind. Stattdessen konzentrierten wir uns auf grundlegende Aspekte, um eine erste explorative Untersuchung ethischer Aspekte aus der Perspektive der Studierenden durchführen zu können.

4 Methodisches Vorgehen

4.1 Befragung und Studienteilnehmende

Eine Online-Umfrage wurde für die Zielsetzung der Studie gewählt, um die Einstellungen und Erfahrungen von Studierenden in Bezug auf KI umfassend zu erforschen. Die Umfrage wurde digital mithilfe der Plattform Qualtrics im Zeitraum von September bis Oktober 2022 durchgeführt. Insgesamt nahmen 636 Studierende der Universitäten St. Gallen und Mannheim teil.

Tabelle 1: Überblick über die Stichprobe

Charakteristik	Absolut	Prozent
Weibliche Studierende	311	49 %
Männliche Studierende	325	51 %
Studierende der Universität St. Gallen	215	34 %
Studierende der Universität Mannheim	421	66 %

An der Universität St. Gallen wurde die Befragung im Rahmen der Vorlesung zur Einführung ins akademische Schreiben für Erstsemesterstudierende durchgeführt. Die Teilnahme an der Befragung war für die Studierenden freiwillig. Insgesamt haben 12.6 % aller Erstsemesterstudierenden von St. Gallen an der Umfrage teilgenommen. Auch an der Universität Mannheim wurden Erstsemesterstudierende verschiedener Fachrichtungen befragt. Die Daten wurden durch Studierende im Rahmen einer Vorlesung zu Grundlagen der Statistik erhoben. Die Studierenden des Kurses haben den Zugangslink zur Online-Befragung erhalten und hatten den Auftrag selbstständig Umfrageteilnehmende an der Universität zu akquirieren. Sie konnten freiwillig auch selbst an der Umfrage teilnehmen. Insgesamt konnten auf diese Art 14.5 % aller Erstsemesterstudierenden von Mannheim motiviert werden, an der Umfrage teilzunehmen. Das Durchschnittsalter aller Befragten beträgt 20.6 Jahre ($SD = 2.5$). Die Stichprobe bezieht sich auf Studierende aus dem deutschsprachigen Raum an zwei Wirtschaftsfakultäten und somit auf Studiengänge in den Fachbereichen Wirtschaft, Recht und Sozialwissenschaften. Unvollständig ausgefüllte Fragebögen wurden aus der untersuchten Stichprobe ausgeschlossen.

4.2 Entwicklung Instrument

Die durchgeführte Befragung besteht neben demographischen Merkmalen aus drei Teilen sowie 37 Items, die für diese Studie ausgewertet wurden (insgesamt umfasst der Fragebogen 135 Items). Die nachfolgende Tabelle gibt zunächst einen Überblick über den Aufbau und Beispiel-Items des Erhebungsinstruments:

Tabelle 2: Aufbau und Beispiel-Items des Erhebungsinstruments (eigene Darstellung)

Untersuchungsgegenstand	Beispiel Items
TEIL I – Auswirkungen von KI	
Chancen von KI	Künstliche Intelligenz (KI) kann unsere Welt verändern. Für wie wahrscheinlich halten Sie die folgenden Vorteile der KI? (5-stufige Likert-Skala) - KI wird die Durchschnittslöhne erhöhen. - KI wird die nationale Sicherheit erhöhen.
Risiken von KI	Künstliche Intelligenz (KI) kann jedoch auch gewisse Nachteile mit sich bringen. Für wie wahrscheinlich halten Sie die folgenden Risiken der KI? (5-stufige Likert-Skala) - KI wird die soziale Ungleichheit vergrößern. - KI wird einigen Menschen zu viel Macht geben.
Risiko- und Chanceneinschätzung von KI für verschiedene Zielgruppen	Nicht jeder wird von der KI auf die gleiche Weise betroffen sein. Halten Sie KI insgesamt für riskant oder vorteilhaft... - Für die demokratische Gesellschaft ist die KI - Für mich persönlich ist die KI
TEIL II – KI-Roboter für Textgenerierung	
Nutzung	Verwenden Sie persönlich Online-Übersetzungstools, wenn Sie fremdsprachige Texte schreiben? (6-stufige Likert-Skala)
Ethische Bedingungen zur Nutzung	In einem akademischen Kontext würde ich die Verwendung von Online-Übersetzungsprogrammen nur dann befürworten, wenn: (Mehrfachnennungen Antworten möglich) - Diese Tools werden nur für Aufgaben verwendet, die nicht benotet werden. - Diese Tools sammeln und speichern keine sensiblen Daten.
TEIL III – KI-Roboter für soziale Interaktionen	
Nutzung	Verwenden Sie persönlich Chatbots? (6-stufige Likert-Skala) - Verwenden Sie persönlich Chatbots bei der Interaktion mit mobilen Geräten? - Nutzen Sie persönlich Chatbots im Bildungsbereich?
Ethische Bedingungen zur Nutzung	Im Bildungsbereich würde ich den Einsatz von Chatbots nur dann befürworten, wenn: (Mehrfachnennungen möglich) - Es wird klar kommuniziert, dass es sich nicht um eine reale Person handelt. - Chatbots haben keine sensiblen Daten gesammelt und gespeichert.

Der Teil I des Fragebogens (Forschungsfrage 1) integriert die Fragebogen-Items aus der Studie von Bao et al. (2022) nach drei Dimensionen:

- (1) Chancen der KI (Durchschnittslöhne erhöhen, nationale Sicherheit erhöhen, individuelle Gesundheit verbessern, Verzerrungen in menschlichen Entscheidungen reduzieren, im Kampf gegen Terrorismus helfen).
- (2) Risiken der KI (die soziale Ungleichheit vergrößern, einigen Menschen zu viel Macht geben, die persönliche Freiheit beschneiden, die Bedeutung des Menschseins verändern, menschliche Arbeitnehmende ersetzen).
- (3) Zielgruppenspezifische Auswirkungen (für die demokratische Gesellschaft, für mich persönlich, für die Studierenden, für Europäer*innen). Verwendet wird wie in der Studie von Bao et al. (2022) eine 5-er Likert-Skala (von 1 = sehr unwahrscheinlich bis 5 = sehr wahrscheinlich bzw. bei Fragekomplex 3, zielgruppenspezifisch von 1 = sehr riskant bis 5 = sehr vorteilhaft).

Teil II des Fragebogens befasst sich mit der Forschungsfrage 2. Die Häufigkeit der Nutzung von KI-Robotern zur Texterstellung wird anhand einer 6-stufigen Likert-Skala erhoben (immer = 6, oft = 5, manchmal = 4, gelegentlich = 3, selten = 2, nie = 1). Die Befragten können angeben, wie häufig sie einen KI-Roboter verwenden oder wie häufig sie eine bestimmte Tätigkeit ausführen. Um den Studierenden ein einheitliches Bild der vier KI-Roboter zur Textgenerierung zu vermitteln, wird ihnen zu Beginn jedes Frageblocks ein Screenshot einer entsprechenden Anwendung gezeigt und die Funktion des jeweiligen KI-Roboters kurz beschrieben (z. B. Erklärung zu maschinellen Übersetzern: Online-Übersetzungshilfen sind Softwaretools (z. B. DeepL, Google Translate), die auf Basis von KI einen Text automatisch in verschiedene Sprachen übersetzen können. Der ethische Einsatz von KI-Robotern wird in Form von Auswahloptionen erhoben. Die Antwortmöglichkeiten beschreiben die verschiedenen Bedingungen, unter denen der Einsatz des jeweiligen KI-Roboters für die Studierenden akzeptabel ist. Die Befragten können für jeden KI-Roboter mehrere Optionen auswählen. Es gibt noch keine validierten Instrumente und Skalen zur Erfassung der ethischen Nutzung. Aufgrund der Komplexität und Nuanciertheit ethischer Überlegungen wurde daher ein explorativer, qualitativer Ansatz mit Antwortoptionen gewählt, um die spezifischen Bedingungen und Szenarien zu identifizieren, die aus Sicht der Studierenden für den Einsatz von KI als wesentlich erachtet werden.

In Teil III (Forschungsfrage 3) liegt der Fokus auf KI-Robotern zur sozialen Interaktion. Analog zu Teil II wird auch hier zu Beginn jedes Frageblocks der entsprechende KI-Roboter kurz beschrieben und ein Screenshot der Applikation oder ein Foto eines passenden physischen Roboters gezeigt. Die Nutzungshäufigkeit von Chatbots wird mit der gleichen 5-stufigen Likert-Skala erfasst. Da virtuelle Avatare und Soziale Roboter zum Zeitpunkt der Befragung noch nicht sehr verbreitet waren, wurde die Nutzungshäufigkeit mit einer dreistufigen Skala erhoben (Skala für digitale Avatare:

Ich habe bereits eine Software, um einen eigenen Avatar zu erstellen. = 1, Ich habe Videos über KI-basierte Avatare gesehen. = 2, Ich habe mich noch nie mit dem Thema ‚digitale Avatare‘ beschäftigt. = 3; Skala für Soziale Roboter: Ich habe bereits mit sozialen Robotern interagiert. = 1, Ich habe Videos über Soziale Roboter gesehen. = 2, Ich bin noch nie mit sozialen Robotern in Kontakt gekommen. = 3). Auch in diesem Teil wurde die ethische Nutzung der drei KI-Robotern (Chatbots, Avatar, Sozialer Roboter) in Form von Auswahloptionen erhoben. Die Studierenden hatten die Möglichkeit, für jeden KI-Roboter Typ mehrere Antwortmöglichkeiten auszuwählen.

5 Ergebnisse

5.1 Interne Konsistenz

Zur Überprüfung der internen Konsistenz der gebildeten Indizes stützen wir uns auf das Cronbachs Alpha (Cronbach, 1951). Aus den Aussagen zur Beurteilung von Chancen und von Risiken wurden zwei Indizes gebildet. Der Chancenindex ($M = 3.2$; $SD = 0.26$; $\alpha = .62$) weist eine eher niedrige interne Konsistenz aus. Dies könnte darauf zurückzuführen sein, dass die Antwortmöglichkeiten der Skala breit gefächert sind und verschiedene Lebensbereiche der Studierenden ansprechen. In der ursprünglichen Studie, aus der die Items übernommen wurden (Bao et al., 2022), war zudem die Stichprobengröße ($N = 2\,700$) erheblich größer. Auch könnten kulturelle Unterschiede zu der in Amerika durchgeführte Studie eine Ursache darstellen (z. B. KI im Kampf gegen Terrorismus). Der Risikoindex ($M = 3.6$; $SD = 0.29$; $\alpha = .73$) weist ein angemessenes Maß an interner Konsistenz auf. Aus den Aussagen über die Einschätzung der Risiken und Chancen von KI für die verschiedenen Zielgruppen lässt sich ein angemessenes Maß an interner Kohärenz ableiten ($M = 3.3$; $SD = 0.25$; $\alpha = .76$).

5.2 Einschätzungen der Auswirkungen von KI

Wie die Befragten die Chancen für positive Auswirkungen von KI in verschiedenen Bereichen einschätzen, zeigt Tabelle 3. Die Prozentangaben beziehen sich auf die Wahrscheinlichkeit, mit der KI bestimmte Auswirkungen erzielen wird. Die Behauptung, KI werde die Durchschnittslöhne steigern, erzielte eine mittlere Bewertung von 2.8, wobei mehr als 40 % der Befragten diese Aussage als eher (sehr) unwahrscheinlich empfinden. Für die nationale Sicherheit und individuelle Gesundheit wurden durchschnittliche Bewertungen von 3.3 bzw. 3.5 abgegeben, wobei ein höherer Grad an Zustimmung zu erkennen war. Die Reduzierung von Verzerrungen in menschlichen Entscheidungen durch KI erhielt eine Bewertung von 3.2, wobei die Meinungen geteilt waren. Hinsichtlich der Hilfe von KI im Kampf gegen Terrorismus wurde eine

durchschnittliche Bewertung von 3.3 abgegeben. Der Chancenindex, der eine Gesamteinschätzung der Studierenden repräsentiert, liegt bei 3.2.

Tabelle 3: Chanceneinschätzung, positive Auswirkungen von KI

Aussage	M (SD)	Wahrscheinlichkeit				
		sehr unwahr- scheinlich	eher unwahr- scheinlich	weder noch	eher wahr- scheinlich	sehr wahr- scheinlich
KI wird die Durchschnittslöhne erhöhen.	2.8 (1.04)	10 %	31 %	32 %	22 %	5 %
KI wird die nationale Sicherheit erhöhen.	3.3 (1.00)	4 %	18 %	28 %	40 %	9 %
KI wird die individuelle Gesundheit verbessern.	3.5 (1.03)	4 %	14 %	24 %	44 %	14 %
KI wird Verzerrungen in menschlichen Entscheidungen reduzieren.	3.2 (1.02)	6 %	18 %	37 %	31 %	8 %
KI wird im Kampf gegen Terrorismus helfen.	3.3 (1.11)	8 %	17 %	28 %	36 %	12 %
Chancenindex	3.2 (1.04)					

Anmerkungen: M = Mittelwert, SD = Standardabweichung

Die Tabelle 4 gibt Einblick in die Risikoeinschätzung der Befragten hinsichtlich der negativen Auswirkungen von KI. Der Durchschnitt der Bewertungen zeigt, dass die Vergrößerung der sozialen Ungleichheit ($M = 3.4$), die Vergabe von zu viel Macht an einige Personen durch KI ($M = 3.9$), die mögliche Beschränkung persönlicher Freiheit ($M = 3.5$), die Veränderung der Bedeutung des Menschseins ($M = 3.4$) und die Befürchtung der Arbeitsplatzersetzung durch KI ($M = 4.0$) als Risiken wahrgenommen werden. Der Risikoindex, der die Gesamteinschätzung repräsentiert, liegt bei 3.6 und ist daher höher als der Chancenindex mit 3.2.

Tabelle 4: Chanceneinschätzung, negative Auswirkungen von KI

Aussage	M (SD)	Wahrscheinlichkeit				
		sehr unwahr- scheinlich	eher unwahr- scheinlich	weder noch	eher wahr- scheinlich	sehr wahr- scheinlich
KI wird die soziale Un- gleichheit vergrößern.	3.4 (0.95)	2 %	14 %	34 %	38 %	12 %
KI wird einigen Menschen zu viel Macht geben.	3.9 (1.00)	3 %	7 %	19 %	42 %	30 %
KI wird die persönliche Freiheit beschneiden.	3.5 (0.97)	3 %	11 %	31 %	40 %	15 %
KI wird die Bedeutung des Menschseins verändern.	3.4 (1.14)	8 %	14 %	27 %	35 %	16 %
KI wird menschliche Ar- beitnehmende ersetzen.	4.0 (1.00)	3 %	14 %	16 %	41 %	34 %
Risikoindex	3.6 (1.01)					

Anmerkungen: M = Mittelwert, SD = Standardabweichung

Die Tabelle 5 zeigt die Risiko- und Chanceneinschätzungen der Befragten von KI für verschiedene Zielgruppen. Im Durchschnitt wird KI in Bezug auf die demokratische Gesellschaft als Risiko betrachtet ($M = 2.9$), wobei 6 % dies als sehr riskant und 4 % als sehr vorteilhaft ansehen. Für sich persönlich sehen die Befragten KI eher als vorteilhaft ($M = 3.3$). Auch für die Gruppe der Studierenden ($M = 3.5$) und generell für Europäer*innen ($M = 3.3$) sehen die Befragten in der KI eher einen Vorteil. Obwohl über die Hälfte der befragten Studierenden die KI als vorteilhaft für Studierende einschätzt, fällt die Chanceneinschätzung für sich persönlich jedoch geringer aus.

Tabelle 5: Risiko- und Chanceneinschätzung von KI für verschiedene Zielgruppen

Aussage	M (SD)	Einschätzung				
		sehr riskant	riskant	weder noch	vorteilhaft	sehr vorteilhaft
Für die demokratische Gesellschaft ist die KI	2.9 (1.00)	6 %	33 %	30 %	28 %	4 %
Für mich persönlich ist die KI	3.3 (0.97)	4 %	17 %	32 %	39 %	8 %
Für die Studierenden ist die KI	3.5 (0.89)	1 %	13 %	26 %	50 %	9 %
Für Europäer*innen ist die KI	3.3 (0.97)	3 %	20 %	32 %	38 %	7 %
Zielgruppenindex	3.3 (0.96)					

Anmerkungen: M = Mittelwert, SD = Standardabweichung

5.3 KI-Roboter Typen mit menschenähnlicher Leistung

Das Nutzungsverhalten der Befragten von KI-gestützter Textgenerierung ist in Tabelle 6 dargestellt. Eine große Mehrheit der Befragten nutzt maschinelle Übersetzer (M = 4.5). Zudem wird das ‚Back-and-Forth‘-Übersetzen, bei dem eigene Texte hin und her übersetzt werden, ebenfalls häufig verwendet (M = 3.4). Schreibassistenten werden manchmal verwendet (M = 2.9). Zusammenfassungsgeneratoren werden am seltensten verwendet, wobei der Mittelwert bei 2.1 liegt.

Tabelle 6: Nutzung KI-Roboter zur Texterstellung

Aussage	M (SD)	Häufigkeit					
		immer	oft	manchmal	gelegentlich	selten	nie
Maschinelle Übersetzer	4.5 (1.19)	19 %	40 %	23 %	8 %	8 %	1 %
Maschinelle Übersetzer: ‚Back-and-Forth‘	3.4 (1.74)	12 %	22 %	18 %	11 %	16 %	22 %
Schreibassistenten	2.9 (1.76)	9 %	15 %	18 %	10 %	15 %	34 %
Umschreibe-/ Zusammenfassungsgeneratoren	2.1 (1.49)	2 %	8 %	14 %	6 %	15 %	55 %

Anmerkungen: M = Mittelwert, SD = Standardabweichung

Auf die explizite Frage nach KI-basierten Textgeneratoren (nicht in der Tabelle enthalten) gaben 82 % der Befragten an, diese noch nie genutzt zu haben. Nur knapp 2 % der Befragten nutzen regelmäßig kostenpflichtige Textgeneratoren. 9 % der Befragten haben kostenpflichtige Textgeneratoren ausprobiert, halten sie aber für zu teuer. 6 % der Befragten finden kostenpflichtige Tools nicht nützlich.

In der nächsten Tabelle 7 wird der Stellenwert der einzelnen Ethikkriterien für KI-Roboter zur Textgenerierung dargestellt. Bei allen definierten KI-Roboter dieser Kategorie wird das Kriterium ‚Gerechtigkeit‘ am häufigsten genannt (M = 46 %). Nur 6 % der befragten Studierenden wären zudem damit einverstanden, dass nur Personen mit einer Rechtschreibschwäche (Legasthenie) ein KI-Roboter zur Textgenerierung verwenden dürfen. Das zweitwichtigste Kriterium (M = 32 %) ist die Gewährleistung des Schutzes der Privatsphäre der Nutzenden. Der transparente Ausweis der Nutzung der KI (Anpassung der Eigenständigkeitserklärung, Markierung der KI-generierten Textpassagen) scheint dagegen weniger relevant zu sein, am wenigsten bei maschinellen Übersetzern mit 10 % der Studierenden. Aber auch bei den KI-gestützten Textgeneratoren sind die Mittelwerte mit 22 % im Vergleich zu den anderen ethischen Aspekten relativ niedrig.

Tabelle 7: Ethische Aspekte zur Nutzung von KI-Robotern zur Texterstellung

KI-Roboter	Transparenz	Gerechtigkeit	Privatsphäre	Kein Schaden
Maschinelle Übersetzer	10 %	56 %	39 %	23 %
Schreibassistenten	18 %	49 %	31 %	17 %
Umschreibe-/ Zusammenfassungs- generatoren	22 %	45 %	32 %	24 %
KI-basierte Textgeneratoren	22 %	35 %	24 %	18 %
Durchschnitt	18 %	46 %	32 %	21 %

5.4 KI-Roboter Typen mit menschenähnlicher Erscheinung für Soziale Interaktionen

Das Nutzungsverhalten von KI-Robotern mit menschenähnlicher Interaktionsform ist in Tabelle 8 abgebildet. Eine deutliche Mehrheit der Befragten (64 %) gibt an, Chatbots selten bis nie zu nutzen (M = 2.3).

Tabelle 8: Nutzung von KI-Robotern zur sozialen Interaktion

KI-Roboter	M (SD)	Häufigkeit					
		immer	oft	manchmal	gelegentlich	selten	nie
Chatbots	2.3 (1.43)	3 %	6 %	16 %	12 %	21 %	43 %

Anmerkungen: M = Mittelwert, SD = Standardabweichung

Insgesamt haben 72 % der Befragten bereits einen persönlichen Avatar erstellt, 23 % haben Videos mit digitalem Avatar gesehen und 5 % hatten noch keine Berührung mit Avataren. 9 % der Befragten haben bereits mit sozialen Robotern interagiert, 30 % haben Videos über soziale Roboter gesehen und 60 % hatten noch keinen Kontakt mit sozialen Robotern.

Tabelle 9 zeigt die Bedeutung der einzelnen ethischen Kriterien für die menschen-ähnliche Interaktion mit KI-Robotern. Die Abweichungen zwischen den durchschnittlichen Werten fallen moderat aus. Dabei erzielt das Kriterium der Transparenz den höchsten Durchschnittswert aller Nennungen (M = 24 %). Zwischen den drei Robotertypen sind keine signifikanten Unterschiede hinsichtlich der ethischen Nutzung festzustellen.

Tabelle 9: Ethische Aspekte zur Nutzung von KI-Robotern zur Texterstellung

KI-Roboter	Transparenz	Gerechtigkeit	Privatsphäre	Kein Schaden
Chatbots	24 %	23 %	20 %	19 %
Avatare	26 %	20 %	14 %	18 %
Soziale Roboter	21 %	19 %	23 %	20 %
Durchschnitt	24 %	21 %	19 %	19 %

6 Diskussion

6.1 Einschätzungen der Auswirkungen von KI

Der ermittelte Chancenindex (3.2) lässt darauf schließen, dass Studierende der Künstlichen Intelligenz gewisse positive Auswirkungen auf die Gesellschaft zuschreiben. Die Daten zeigen, dass Studierende gemischte Meinungen zur Rolle der KI in verschiedenen Bereichen haben. Eine deutliche Mehrheit glaubt, dass KI die individuelle Gesundheit verbessern wird, was ein hohes Vertrauen in die positiven Auswirkungen der KI auf den Gesundheitssektor widerspiegelt. Auch sieht die Mehrheit der Studierenden KI positiv im Kontext der nationalen Sicherheit und im Kampf gegen Terrorismus. Bei der Frage, ob KI-Verzerrungen in menschlichen Entscheidungen reduzieren kann, sind die Meinungen unter den Studierenden

gespalten. Dies deutet darauf hin, dass die Studierenden sowohl das Potenzial als auch die Grenzen der KI in der Entscheidungsfindung erkennen. Nur eine Minderheit der Studierenden hält es für wahrscheinlich, dass KI die Durchschnittslöhne erhöhen wird.

Die wahrgenommenen Risiken übersteigen die erkannten Chancen (Risikoindex 3.6). Dreiviertel der Studierenden sind der Meinung, dass KI menschliche Arbeitskräfte ersetzen wird, was auf eine weit verbreitete Ansicht hindeutet, dass KI zu signifikantem Arbeitsplatzverlust führen könnte. Auch befürchtet eine deutliche Mehrheit, dass KI bestimmten Personenkreisen zu viel Macht einräumen könnte, was Bedenken bezüglich einer potenziellen Machtkonzentration, möglicherweise insbesondere durch große Technologiekonzerne, widerspiegelt. Die Einschätzungen verdeutlichen darüber hinaus, wie tiefgreifend die Studierenden die potenziellen Auswirkungen der KI auf individuelle Freiheiten und das menschliche Selbstverständnis einschätzen. Dies könnte in Datenschutzbedenken oder in einer vermehrten Nutzung von KI in Überwachungssystemen begründet liegen. Aber auch dass die menschliche Autonomie durch KI-Systeme untergraben oder zunehmende Manipulation von Meinungen und Verhalten durch die Nutzung von KI in der Informationsverbreitung und Werbung führen könnte. Die wachsende Integration von KI in den Alltag könnte nach Ansicht vieler Studierenden die Mensch-Maschine-Beziehung verändern und beeinflusst folglich, wie Menschen ihre Rolle gegenüber Technologien sehen und was wiederum das berufliche Selbstverständnis der Menschen tiefgreifend verändern könnte. Zudem könnte KI traditionelle Auffassungen von menschlicher Intelligenz, Kreativität und Einzigartigkeit infrage stellen, indem sie Aufgaben übernimmt, die bisher als spezifisch menschlich galten.

Spezifisch nach den Zielgruppen gefragt, zeigt sich, dass die Einschätzung der Auswirkungen von KI je nach Kontext und persönlicher Betroffenheit variieren kann, wobei die unmittelbaren Vorteile für Studierende stärker wahrgenommen werden als die komplexeren gesellschaftlichen Folgen in einer Demokratie. Studierende könnten direkte Vorteile von KI im Studium sehen, wie personalisiertes Lernen, effizientere Forschungswerkzeuge und verbesserte Karrieremöglichkeiten. Allerdings ist die Einschätzung der persönlichen Chancen durch KI geringer im Vergleich zu anderen Studierenden. Diese Diskrepanz könnte darauf zurückzuführen sein, dass einige Befragte eine ‚Nicht für mich, aber gut für andere‘-Haltung gegenüber KI in der Hochschullehre einnehmen. Dies könnte einerseits auf ein erhöhtes Selbstwirksamkeitsgefühl hindeuten, bei dem sich Individuen als kompetenter oder weniger auf externe Hilfe angewiesen betrachten im Vergleich zu anderen. Andererseits könnte es aber auch darin begründet liegen, dass sich Studierende im Vergleich zu anderen unterlegen fühlen oder ein Mangel an Wissen über die konkreten Anwendungspotenziale von KI wahrnehmen.

6.2 Nutzungserfahrung mit KI-Roboter Typen

Die Analyse der Daten legt nahe, dass die Nutzung der befragten KI-Roboter Typen durch Studierende insgesamt als moderat einzustufen ist vor dem Erscheinen von ChatGPT. Diese geringe Nutzung könnte auf unzureichende Kenntnisse über die Vorteile der KI, mangelnde Fähigkeiten in ihrer Bedienung oder eingeschränkten Zugang zurückzuführen sein. Insbesondere KI-Roboter zur Inhaltsgenerierung wurden als kostenpflichtig und nach Meinung der Studierenden als noch nicht ausreichend leistungsfähig empfunden. Ein interessanter Befund ist es, dass bereits erstaunlich viele Studierende die Übersetzungssoftware DeepL oft bzw. sogar immer (12 % der Befragten) für das Schreiben und Verbessern von Texten in der eigenen Sprache („Back-and-Forth Translation“) benutzt haben. Dies könnte ein Indiz dafür sein, dass sich die Schreibpraktiken der Studierenden durch die Verwendung von KI schon vor der Einführung von ChatGPT in einem drastischen Veränderungsprozess befanden.

Bei der Nutzung von KI-Robotern für soziale Interaktionen zeigen die Ergebnisse, dass die Nutzung von Chatbots direkt vor dem Erscheinen von ChatGPT im November 2022 noch nicht sehr verbreitet war. Dahingegen war die Erstellung eigener Avatare bereits erstaunlich hoch. Dies könnte darauf hinweisen, dass sich Avatar-basierte Dialogschnittstellen künftig schnell verbreiten könnten, da eine grundsätzliche Akzeptanz vorhanden zu sein scheint. Wie zu erwarten, hat nur ein sehr kleiner Teil der Befragten bereits mit einem sozialen, physischen Roboter interagiert.

6.3 Ethische Kriterien für die Nutzung von KI-Roboter Typen

Die Analyse der ethischen Einstellungen zeigt, dass eine generelle Ablehnung der KI-Systeme aus ethischen Gründen nicht vorliegt. Die erhobenen Daten zeigen jedoch ein gemischtes Bild hinsichtlich der Bedeutung ethischer Nutzungsaspekte. Es fällt auf, dass viele Befragte den Einsatz bestimmter KI-Systeme uneingeschränkt unterstützen würden. Es könnte sein, dass den Studierenden die potenziellen Nachteile nicht bewusst sind, wenn wichtige ethische Aspekte bei der Nutzung von KI außer Acht gelassen werden. Eine andere Deutung könnte sein, dass die Studierenden sich der Risiken bewusst sind, diese aber durch den erhofften Nutzen der KI ausgeglichen sehen (Green, 2021).

Die Zustimmung zu ethischen Grundprinzipien variiert je nach Art des KI-Roboter Typs. Studierende messen bei KI-Robotern zur Texterstellung der Gerechtigkeit mehr Bedeutung bei, während bei KI-Robotern für soziale Interaktionen die Transparenz als wichtiger angesehen wird. Das starke Interesse der Studierenden an einem gerechten Zugang zu bestimmten KI-Systemen könnte auf dem Wunsch basieren, allen gleiche Chancen zu bieten, oder darauf, nicht benachteiligt zu werden.

Die geringe Zustimmung zur Bereitstellung der KI für Studierende mit Lese-Rechtschreibschwäche könnte auf die Befürchtung von Benachteiligungen hindeuten.

Der höhere Zuspruch nach Transparenz bei sozialen Interaktionen könnte darin begründet liegen, dass Menschen nicht unwissentlich mit einem KI-Roboter kommunizieren und ggf. von einer Maschine beeinflusst oder manipuliert werden sollten. Bei der Erstellung von Texten mittels KI wird hingegen nur von wenigen Studierenden die Transparenz als ethisches Nutzungsprinzip ausgewählt. Nur eine Minderheit unterstützen es, dass die mit KI erstellten Passagen deutlich als solche gekennzeichnet werden, die Eigenständigkeitserklärung am Ende einer schriftlichen Arbeit angepasst und die Verwendung solcher Hilfsmittel spezifiziert werden sollte. Dies könnte ein Mangel an ethischem Bewusstsein hinsichtlich der Integrität wissenschaftlicher Arbeiten und der Bedeutung von Eigenständigkeit beinhalten (Cotton et al., 2023). Viele könnten KI als bloßes Werkzeug betrachten, ähnlich einem Taschenrechner oder Textverarbeitungsprogramm, das keiner speziellen Kennzeichnung bedarf (Lodge, Yang, Furze & Dawson, 2023). Zum Erhebungszeitpunkt der Umfrage im September 2022 gab es generell an Universitäten noch keine klaren Richtlinien oder Normen zur Verwendung von KI in akademischen Arbeiten, was zu Unsicherheit und mangelnder Unterstützung für solche Maßnahmen führen könnte. Studierende könnten zudem befürchten, dass die Offenlegung der Verwendung von KI in ihren Arbeiten zu einer negativen Bewertung oder einem Vertrauensverlust seitens der Lehrenden führen könnte.

Überraschenderweise wird der Schutz der Privatsphäre und des Datenschutzes von den Studierenden bei KI-Robotern generell gering priorisiert. Die Aspekte wählten sie eher noch bei KI-Robotern mit menschenähnlicher Leistung ($M = 32\%$) als bei menschenähnlicher Interaktion ($M = 19\%$) aus. KI-Roboter zur menschenähnlichen Interaktion könnten allerdings auch Datenschutzbedenken hervorrufen, da sie möglicherweise als invasiver empfunden werden. Insbesondere können soziale Roboter durch ihre Sensorik multimodale Daten ihrer Umwelt aufnehmen (Murphy, 2019). Die Studierenden haben möglicherweise ein begrenztes Verständnis und Erfahrungswerte für die Implikationen von Datenschutz und Privatsphäre im Kontext von KI.

Schadensprävention und Solidarität sind für die ethische Nutzung von KI-Robotern nicht sehr bedeutsam für Studierende. Bemerkenswert ist der Befund, dass Studierende zwar das Risiko des Verlustes von Arbeitsplätzen durch KI als hoch bewerten, dennoch den Einsatz von KI in der Hochschullehre nicht ausschließlich dann unterstützen, wenn dabei keine menschlichen Arbeitsplätze ersetzt werden. Dies könnte darin begründet liegen, dass Studierende diesen Zusammenhang nicht erkennen bzw. sich zu wenig in einer entscheidungs- und gestaltungsorientierten Rolle in ihrem Umfeld sehen (Morley et al., 2023). Des Weiteren könnten die Studierenden

persönliche Vorteile durch den Einsatz von KI in ihrem Studium oder ihrer zukünftigen Karriere sehen, die sie gegenüber den potenziellen Risiken für den allgemeinen Arbeitsmarkt priorisieren. Einige Studierende könnten zudem der Ansicht sein, dass der technologische Fortschritt und die damit verbundene Automatisierung durch KI unvermeidlich sind. Sie befürworten daher den Einsatz von KI, auch wenn damit Arbeitsplatzverluste einhergehen könnten, da sie dies als Teil des natürlichen technologischen Wandels sehen.

Ein wichtiger Aspekt, welcher in der Interpretation der Ergebnisse beachtet werden sollte, ist der Erhebungszeitraum, welcher kurz vor der Einführung von ChatGPT liegt. Daher muss davon ausgegangen werden, dass KI-Roboter zum Zeitpunkt der Erhebung für Studierende lediglich eine untergeordnete Rolle spielten und entsprechend noch keine ausgeprägten Überzeugungen im Hinblick auf die ethischen Voraussetzungen zum Einsatz von KI in der Hochschulbildung vorhanden waren. Die Erhebung der Daten kurz vor Erscheinen von ChatGPT birgt zudem die Limitation, dass sich die Studie auf ein Verständnis von KI und Machine Learning bezieht, das darauf ausgerichtet ist, bestehende Daten zu verstehen, Muster zu erkennen und daraus Schlüsse zu ziehen, z. B. für die Datenanalyse, Vorhersage von Kundenverhalten, Erkennung von Betrug oder für die Automatisierung von Routineaufgaben. Im Unterschied dazu stellen die neuen technologischen Entwicklungen auf der Basis großer Sprachmodelle wie z. B. ChatGPT eine generative KI dar, die darauf abzielt, auf Basis dieser Verständnisse neue, eigenständige Inhalte zu erschaffen. Mit diesen technologischen Entwicklungen gehen somit auch neue Chancen, wie z. B. die effiziente Unterstützung bei kreativen Prozessen, sowie neue Risiken einher, wie z. B. Datenhalluzinationen, wenn KI-Inhalte generiert, die zwar kohärent und plausibel erscheinen können, aber nicht auf Fakten basieren oder sogar völlig fiktiv sind.

Zusätzliche Einschränkungen dieser Studie ergeben sich aus der Zusammensetzung der Stichprobe und der Qualität des verwendeten Instruments. Die Stichprobe beschränkt sich auf Studierende aus dem deutschsprachigen Raum an zwei Wirtschaftsfakultäten, also auf die Fachbereiche Wirtschaft, Recht und Sozialwissenschaften. Die Stichprobengröße ist zudem deutlich kleiner als in der Chancen-Risiko-Studie von Bao et al. (2022). Hinsichtlich der Qualität des Instruments weist der Chancen-Index eine mittlere interne Konsistenz auf, die niedriger ist als in der Studie von Bao et al. (2022), was möglicherweise auf kulturelle Unterschiede und eine wesentlich kleinere Stichprobengröße zurückzuführen ist.

7 Implikationen und Ausblick

Die Ergebnisse der Studie deuten darauf hin, dass es entscheidend ist, die Perspektiven der Studierenden bei der Einführung von KI in der Hochschullehre stärker zu beachten. Dies umfasst insbesondere Maßnahmen zur Förderung der Kompetenzen

und die Einbindung von Möglichkeiten für kritische Diskussionen über die Auswirkungen von KI (Laupichler, Aster, Schirch & Raupach, 2022). Es ist daher von Bedeutung, eine interdisziplinäre KI-Kompetenz zu entwickeln, die über verschiedene Studiengänge hinweg gefördert werden sollte, um sowohl die Möglichkeiten als auch die Herausforderungen, die durch KI entstehen, zu verstehen und verantwortungsvoll einzusetzen (Ng, Leung, Chu & Qiao, 2021; Zhao, Pinto Llorente & Sánchez Gómez, 2021). In zukünftigen Forschungsprojekten sollte der Chancen- und Risikoindex weiterentwickelt werden, um die Entwicklungen im Bereich der generativen KI stärker zu berücksichtigen.

Darüber hinaus sind Modelle und Entwicklungsansätze für die Förderung von KI-Kompetenzen in der Hochschullehre zu entwickeln und deren Wirksamkeit zu erforschen (Laupichler et al., 2022; Zhao et al., 2021). Normen und Richtlinien für die ethische Nutzung von KI-Robotern für die Textgenerierung wie ChatGPT zu entwickeln, stellt derzeit eine bedeutende und komplexe Herausforderung für Hochschulen dar (Lo, 2023). Die Kennzeichnungspflicht von KI-generierten Inhalten in akademischen Arbeiten kann dazu beitragen, die ethischen, akademischen und pädagogischen Standards zu stärken und aufrechtzuerhalten. Eine klare Kennzeichnung trägt zur Wahrung der akademischen Integrität bei. Die Kennzeichnung unterstützt die Einhaltung ethischer Standards in der akademischen Arbeit, indem sie die Eigenleistung der Studierenden von der maschinellen Erstellung abgrenzt. Lehrende können die Qualität der KI-generierten Inhalte besser beurteilen und einschätzen, wie gut Studierende diese KI-Systeme nutzen und verstehen. Diese Maßnahme könnte dazu führen, dass das Bewusstsein der Studierenden für einen verantwortungsvollen Umgang mit KI und deren Auswirkungen auf ihre Lernprozesse gefördert werden kann (Zhai, 2022). Die Ergebnisse unserer Studien lassen jedoch darauf schließen, dass die Akzeptanz der Studierenden gegenüber dieser Maßnahme eher gering sein dürfte. In weiteren Forschungsarbeiten möchten wir daher diese Studierendenperspektive genauer erforschen, um nach Lösungen zu suchen, die einen ethisch verantwortungsbewussten Einsatz von KI in der Hochschulbildung ermöglichen und gleichzeitig die Entwicklung erforderlicher KI-Kompetenzen unterstützen, anstatt sie zu behindern.

Literatur

- Alshami, A., Elsayed, M., Ali, E., Eltoukhy, A. E. E. & Zayed, T. (2023). Harnessing the Power of ChatGPT for Automating Systematic Review Process: Methodology, Case Study, Limitations, and Future Directions. *Systems*, 11 (7), 351. <https://www.doi.org/10.1080/10494820.2023.2253858>
- Bao, L., Krause, N. M., Calice, M. N., Scheufele, D. A., Wirz, C. D., Brossard, D., Newman, T. P. & Xenos, M. A. (2022). Whose AI? How different publics think about AI and its social impacts. *Computers in Human Behavior*, 130, 107182. <https://www.doi.org/10.1016/j.chb.2022.107182>
- Boyd-Graber, J., Okazaki, N. & Rogers, A. (2023). ACL 2023 policy on AI writing assistance. Available under: <https://2023.aclweb.org/blog/ACL-2023-policy/> [15.11.2023].

- Cotton, D. R. E., Cotton, P. A. & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 1-12. <https://www.doi.org/10.1080/14703297.2023.2190148>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16 (3), 297-334.
- Dang, J. & Liu, L. (2022). Implicit theories of the human mind predict competitive and cooperative responses to AI robots. *Computers in Human Behavior*, 134, 107300. <https://www.doi.org/10.1016/j.chb.2022.107300>
- Europäische Kommission (2022). Ethische Leitlinien für Lehrkräfte über die Nutzung von KI und Daten für Lehr- und Lernzwecke. Verfügbar unter: https://www.raonline.ch/pages/edu/pdf/CT/EU_KI-Leitlinien_Schulen2022b.pdf [07.12.2023].
- Floridi, L. & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1 (1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Frymier, A. B., Shulman, G. M. & Houser, M. (1996). The development of a learner empowerment measure 1. *Communication Education*, 45 (3), 181-199. <https://www.doi.org/10.1080/03634529609379048>
- Garrel, J. von, Mayer, J. & Mühlfeld, M. (2023). Künstliche Intelligenz im Studium - Eine quantitative Befragung von Studierenden zur Nutzung von ChatGPT & Co. Verfügbar unter: https://opus4.kobv.de/opus4-h-da/frontdoor/deliver/index/docId/395/file/befragung_ki-im-studium.pdf [07.12.2023].
- Green, B. (2021). The Contestation of Tech Ethics: A Sociotechnical Approach to Technology Ethics in Practice. *Journal of Social Computing*, 2 (3), 209-225. <https://www.doi.org/10.23919/JSC.2021.0018>
- Guggemos, J., Seufert, S., Sonderegger, S. & Burkhard, M. (2022). Social robots in education: Conceptual overview and case study of use. In D. Ifenthaler, P. Isaias & D. G. Sampson (Eds.), *Orchestration of learning environments in the digital world* (pp. 173-195). Cham: Springer International Publishing.
- Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1 (9), 389-399. <https://www.doi.org/10.1038/s42256-019-0088-2>
- Kaeser, E. (2022). Chat GPT: Der Roboter schreibt nicht, er schwafelt: Neun Missverständnisse um die Künstliche Intelligenz. *NZZ*. Verfügbar unter: <https://www.nzz.ch/technologie/chat-gpt-der-roboter-schreibt-nicht-er-schwafelt-id.1717513> [07.12.2023].
- Kim, S. (2022). Working with robots: Human resource development considerations in human-robot interaction. *Human Resource Development Review*, 21 (1), 48-74. <https://www.doi.org/10.1177/15344843211068810>
- Kong, S.-C., Man-Yin Cheung, W. & Zhang, G. (2021). Evaluation of an artificial intelligence literacy course for university students with diverse study backgrounds. *Computers and Education: Artificial Intelligence*, 2, 100026. <https://www.doi.org/10.1016/j.caeai.2021.100026>
- Laupichler, M. C., Aster, A., Schirch, J. & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. *Computers and Education: Artificial Intelligence*, 3, 100101. <https://www.doi.org/10.1016/j.caeai.2022.100101>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I. & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education*, 21 (2), 100790. <https://www.doi.org/10.1016/j.ijme.2023.100790>
- Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13 (4), 410. <https://www.doi.org/10.3390/educsci13040410>
- Lodge, J. M., Yang, S., Furze, L. & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? *Learning: Research and Practice*, 9 (2), 117-124. <https://www.doi.org/10.1080/23735082.2023.2261106>
- Minge, M. (2012). Modalitätenwahl bei multimodaler Interaktion. Verfügbar unter: https://www.researchgate.net/publication/263714313_Modalitätenwahl_bei_multimodaler_Interaktion [07.12.2023].
- Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M. & Floridi, L. (2023). Operationalising AI ethics: barriers, enablers and next steps. *AI & SOCIETY*, 38 (1), 411-423. <https://doi.org/10.1007/s00146-021-01308-8>
- Murphy, R. (2019). *Introduction to ai robotics* (2nd Ed.). Cambridge: The MIT Press.

- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://www.doi.org/10.1016/j.caeai.2021.100041>
- Roesler, E., Manzey, D. & Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science robotics*, 6 (58), eabj5425. <https://www.doi.org/10.1126/scirobotics.abj5425>
- Wang, X., Pang, H., Wallace, M. P., Wang, Q. & Chen, W. (2022). Learners' perceived AI presences in AI-supported language learning: a study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning*, 1-27. <https://www.doi.org/10.1080/09588221.2022.2056203>
- Weidemann, B. (2002). *Informationen und Lernen mit Multimedia und Internet*. Weinheim: Verlagsgruppe Beltz.
- Winkler, R., Söllner, M. & Leimeister, J. M. (2021). Enhancing problem-solving skills with smart personal assistant technology. *Computers & Education*, 165, 104148. <https://www.doi.org/10.1016/j.compedu.2021.104148>
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M. & Drachsler, H. (2021). Are we there yet? - a systematic literature review on chatbots in education. *Frontiers in artificial intelligence*, 4, 654924. <https://www.doi.org/10.3389/frai.2021.654924>
- Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN Electronic Journal*, 1-18. <https://www.doi.org/10.2139/ssrn.4312418>
- Zhao, Y., Pinto Llorente, A. M. & Sánchez Gómez, M. C. (2021). Digital competence in higher education research: a systematic literature review. *Computers & Education*, 168, 104212. <https://www.doi.org/10.1016/j.compedu.2021.104212>

Autor*innen

Prof. Dr. Sabine Seufert, Universität St. Gallen, Direktorin des Instituts für Bildungsmanagement und Bildungstechnologien

Lukas Spirgi, Universität St. Gallen, Institut für Bildungsmanagement und Bildungstechnologien

Dr. Jan Delcker, Universität Mannheim, Lehrstuhl für Wirtschaftspädagogik – Learning, Design & Technology

Joana Heil, Universität Mannheim, Lehrstuhl für Wirtschaftspädagogik – Learning, Design & Technology

Prof. Dr. Dirk Ifenthaler, Universität Mannheim, Lehrstuhlinhaber für Wirtschaftspädagogik – Technologiebasiertes Instruktionsdesign

Korrespondenz an: sabine.seufert@unisg.ch

Josef Guggemos, Jacqueline Schmidt & Roland Happ

Eine Frage von Macht: Einstellungen angehender Lehrkräfte zu den ethischen Grundsätzen des Einsatzes künstlicher Intelligenz in der Bildung

Die Integration künstlicher Intelligenz (KI) in die Bildung hat das Potenzial, Lernprozesse zu verbessern. Bei einer KI-Implementierung sind auch ethische Aspekte zu berücksichtigen. In dieser Studie werden die Einstellungen (als Summe an Überzeugungen) angehender Lehrkräfte zu den ethischen Grundsätzen des Einsatzes von KI im Bildungswesen untersucht. Das sind insbesondere Transparenz, Gerechtigkeit und Fairness, Nichtschädlichkeit und Wohlbefinden, Verantwortung, Privatsphäre, Autonomie und Macht über Menschen. Die Stichprobe umfasst 90 angehende Lehrkräfte aus dem berufsbildenden Bereich. Die Ergebnisse zeigen, dass das Messmodell zur Bewertung der Einstellung von Lehrkräften zu den ethischen Grundsätzen der KI-Nutzung reliabel und valide ist. Eine latente Profilanalyse offenbart drei Profile von angehenden Lehrkräften mit jeweils unterschiedlichen Einstellungsstrukturen. Die Lehrkräfte unterscheiden sich vor allem hinsichtlich ihrer Einschätzung, ob es KI erlaubt sein sollte, harte Macht über Lehrkräfte und Lernende auszuüben. Es lässt sich kein Zusammenhang zwischen Profildugehörigkeit und Kontextvariablen, z. B. dem Geschlecht, feststellen. Die Ergebnisse sind von Bedeutung für den Prozess der Integration von KI in den Unterricht und für die Entwicklung von Fortbildungsmaßnahmen, die auf die Einstellungsprofile zugeschnitten werden können.

Schlagwörter: Einstellungen – Ethik – Künstliche Intelligenz – Latente Profilanalyse – Lehrkräftebildung

A matter of power: prospective teachers' attitudes towards the ethical principles of artificial intelligence use in education

The integration of artificial intelligence (AI) in education holds potential for improving learning processes. However, ethical aspects surrounding AI implementation have to be considered. This study explores prospective teachers' attitudes (as a sum of beliefs) towards the ethical principles of AI use in education, namely, transparency, justice and fairness, non-maleficence and well-being, responsibility, privacy, autonomy, and power over people. The sample comprises 90 prospective teachers from vocational education and training. The results indicate that our measurement model for assessing teachers' attitudes towards the ethical principles of AI use is reliable and valid. Latent profile analysis reveals three distinct profiles of prospective teachers with a different attitude structure. The teachers mainly differ in their belief whether AI should be allowed to exert hard power over teachers and students. We cannot find any association between profile membership and context variables, e. g., gender. The results have implications for the process of integrating AI in education and the development of instructional measures tailored to specific attitude profiles.

Keywords: Artificial intelligence – attitudes – ethics – latent profile analysis – teacher education

1 Introduction

Artificial intelligence (AI) has the potential to improve educational processes (Chen, Chen & Lin, 2020; Hwang, Xie, Wah & Gašević, 2020). The term AI is often used and defined inconsistently (André et al., 2021; Seufert, Guggemos & Ifenthaler, 2021a). AI generally refers to systems that automate intellectual activities to a certain extent (e. g., Peissner, Kötter & Zaiser, 2019). In public discourse, AI is often reduced to machine learning (Holmes, Bialik & Fadel, 2019; Seufert et al., 2021a), which is a sub-area of AI. Examples of AI-based educational technology are intelligent tutorial systems, automated essay scoring systems, chatbots, and social robots (Belpaeme, Kennedy, Ramachandran, Scassellati & Tanaka, 2018; Kasneci et al., 2023; Zawacki-Richter, Marín, Bond & Gouverneur, 2019).

In order to be successful, the use of technology in education, and consequentially AI, has to follow theories about how people learn (Sweller, 2020; Zawacki-Richter et al., 2019). Besides this, technical and ethical constraints have to be considered (Guggemos, Seufert, Sonderegger & Burkhard, 2022a). Technical constraints imply that desired activities cannot be realized at the current technical state of the art. Currently available AI technology can be referred to as weak (narrow) AI (Wang & Siau, 2019), i. e., its capabilities are restricted to specific tasks. For instance, an intelligent tutorial system can cognitively activate students, but it cannot manage classroom behavior; social robots are restricted in their ability to navigate in a naturalistic classroom setting. Moreover, ethical constraints are important. Organizations that do not follow ethical principles may lose their moral legitimacy (Suchman, 1995). For instance, data leakage can compromise the privacy of the users and lead to reputation damage for the accountable organization (Harel, Shabtai, Rokach & Elovici, 2010). The same may be the case for an organization that uses AI systems that makes unfair decisions (Marcinkowski, Kieslich, Starke & Lünich, 2020). Literature reviews have been performed about ethical issues when using AI in education (Akgun & Greenhow, 2022; Holmes et al., 2022). Their conclusion may be that AI, on the one hand, can yield substantial benefits for education but, on the other, faces severe challenges (Akgun & Greenhow, 2022). Although these literature reviews provide an overview of important ethical principles for the use of AI in education, they do not offer empirical evidence about how these principles are perceived by crucial stakeholders in education.

In the European Union, the High-Level Expert Group on Artificial Intelligence (2019) has developed a framework for the ethical use of AI. It comprises standards for an ethical AI. This framework has been adapted for the education and training system (European Commission, 2022). The seven standards in this framework are consistent with the influential literature review of Jobin, Ienca and Vayena (2019). This literature review reports a convergence around five key ethical principles, namely, trans-

parency, justice and fairness, non-maleficence, responsibility, and privacy. A more recent literature review also names these five principles as central in education (Adams, Pente, Lemermeyer & Rockwell, 2023).

When it comes to technology integration into education, teachers play a key role (Seufert, Guggemos & Sailer, 2021b). Technology acceptance models, as well as the skill-will-tool model, highlight the important role of teachers' attitudes in this process (Knezek & Christensen, 2016; Scherer, Siddiq & Tondeur, 2019). In the course of the integration of AI-based technology into education, new challenges have emerged, e. g., concerning bias and privacy (Pargmann & Berding, 2022). There are only a few studies that attempt to analyze the attitudes and beliefs of (prospective) teachers towards AI tools. In particular, the perceived benefits and potential of the technology are addressed (an overview is provided by Schmidt, 2024, pp. 71-73). Teachers' attitudes, particularly towards the ethical challenges of AI, have, to our knowledge, received little attention in this context so far (e. g., Lindner & Berges, 2020). The available literature reviews discuss pertinent ethical principles when using AI (in education); however, they do not address how these principles are regarded by teachers. This may be an issue as teachers can be regarded as gatekeepers when it comes to the use of educational technology. An empirical study may be helpful to shed light on teachers' attitudes towards the ethical principles of AI use in education.

Attitudes can be conceptualized as a sum of beliefs (Instefjord & Munthe, 2017). Ethical beliefs address the perception of how something should be (done) (Heller & Phillips, 2020). An example could be the belief that AI-based technology should treat all students fairly. Beliefs may not necessarily be informed by sound knowledge. Pega (2018) reports that adults around the globe think they have a fair knowledge of AI. However, at the same time, they were unable to answer basic questions about AI. Consequentially, their beliefs about and attitudes towards AI may be biased. If someone lacks knowledge about the capabilities of AI at the current technical state of the art, they may form beliefs about AI based on the media; however, portrayals of AI in the media are often biased (Long & Magerko, 2020). Overall, misinformed attitudes towards AI may evolve from those beliefs. Depending on the media content consumed, attitudes can be overly positive or overly negative. For instance, a teacher may refuse to use AI due to worries about such technology taking over full control in the classroom.

Referring to an overall teacher attitude towards AI is likely to be oversimplified. Rather, there may be latent profiles of teachers that differ in level and structure (Scherer, Howard, Tondeur & Siddiq, 2021). Obtaining knowledge about such latent profiles is beneficial (Hofmans, Wille & Schreuers, 2020; Morin & Marsh, 2015) as they may show that teachers' attitudes towards the ethical principles of AI are heterogeneous and, at the same time, shed light on this heterogeneity. Knowledge

about the different types of profiles may help teacher trainers to consider these attitudes in their instruction. There may be a substantial proportion of prospective teachers who are critical towards AI in education. Such critical views can be discussed to help (prospective) teachers to develop well-informed attitudes regarding the ethical principles of AI in education. Having identified latent attitude profiles, we aim to check if latent profile membership is associated with context variables in an expected way; this allows for a better understanding of the identified latent profiles (Scherer et al., 2021).

Overall, there may be three important research desiderata: first, to operationalize the identified ethical aspects of AI use in education; second, to identify profiles of teachers with similar attitudes (as a sum of beliefs) towards the ethical principles of AI in education; third, to investigate whether there is an association between relevant variables and profile membership. By addressing these research desiderata, we aim to contribute to a better understanding about the ethical aspects of AI in education, from a teacher perspective.

2 Conceptual Basis

2.1 Ethical principles of AI use in education

In the following section, we will present some important ethical principles of AI in education. The selection is based on the literature review of Jobin et al. (2019) concerning the ethical principles of AI in general, and the ethical guidelines on the use of AI in education (EGAI) of the European Commission (European Commission, 2022). Moreover, we consider power over people as an important ethical aspect in education (Niesche & Haase, 2012). Overall, we address transparency (2.1.1), justice and fairness (2.1.2), non-maleficence and beneficence (2.1.3), responsibility and accountability (2.1.4), privacy (2.1.5), and autonomy (2.1.6), as well as power over people (2.1.7). Table 1 provides a synopsis of the principles identified by a review of the literature, and their relationship to the principles addressed in our study.

2.1.1 Transparency

Transparency implies the disclosure of information relevant to stakeholders. Ensuring the reproducibility of results can promote transparency (Haibe-Kains et al., 2020). In terms of AI, this means making the entire machine-learning process available. This includes revealing data, metadata, and the program code (Peng, 2011). In particular, data and the rights to it are seen as an important ethical aspect (Holmes et al., 2022). Furthermore, transparency includes disclosing the use of AI and naming a responsible person or institution (Jobin et al., 2019). Transparency has a very positive connotation in public discourse (Felzmann, Fosch-Villaronga, Lutz & Tamò-

Larrieux, 2020, pp. 3336-3339); however, transparency can also have negative consequences (Felzmann et al., 2020, pp. 3339-3341): data can be confidential and thus not suitable for disclosure. The EGAIE emphasize the importance of transparency for the ethical use of AI in education (European Commission, 2022).

2.1.2 Justice and fairness

Justice is often expressed through fairness (Jobin et al., 2019). Fairness can be considered as the effort to treat and value all individuals and groups equally, regardless of their identity, gender, race or ethnicity, social class, religion, or other personal characteristics (AERA, APA & NCME, 2014, p. 50). Fairness plays an important role in human interaction. In terms of AI, it seems problematic to ascribe a higher level of fairness to such technology than to humans (Mittelstadt, Allo, Taddeo, Wachter & Floridi, 2016). For example, it could be argued that AI is fairer than humans because AI is not subject to the known human errors of judgement (Kahneman & Tversky, 1996). Empirically, this assertion cannot be supported. Rather, AI-based decisions and actions can also be unfair (Tambe, Cappelli & Yakubovich, 2019). This partially arises from the implicit or explicit stereotypes and behavioral patterns acquired in the training process, which are reproduced by the human's initial assessment of the training data (Pargmann & Berding, 2022). Justice does not imply equal treatment of all individuals. Following Rawls' theory of justice (Rawls, 1999), special attention would have to be paid to particularly disadvantaged human beings. The EGAIE also highlight this point by referring to people with disabilities and those with special educational needs (European Commission, 2022).

2.1.3 Non-maleficence and beneficence

Non-maleficence implies the absence of harm (O'Donoghue, 2023, p. 128). Examples are hacking (violation of privacy) and discrimination (lack of justice and fairness) (Jobin et al., 2019). On the positive side, AI would foster the well-being of human beings (Jobin et al., 2019; Smakman, Vogt & Konijn, 2021). This principle of AI being beneficial to humanity is one of the cornerstones of AI principles (Floridi et al., 2018). AI may support people in reaching their goals. Overall, non-maleficence and well-being are prominent ethical principles (European Commission, 2022; Jobin et al., 2019).

Table 1: Ethical principles addressed in the study at hand and their relation to literature reviews and the ethical guidelines on the use of AI in teaching and learning for educators.

Ethical principles in our study	Principles identified by Jobin et al. (2019)	Principles identified by Adams et al. (2023)	Principles of the European Commission (2022)
Transparency	Transparency	Transparency	Transparency
Justice and fairness	Justice and fairness	Justice and fairness	Diversity, non-discrimination and fairness
Non-maleficence and beneficence	Non-maleficence; beneficence	Non-maleficence; beneficence; teachers' well-being	Technical robustness and safety; societal and environmental well-being
Responsibility and accountability	Responsibility	Responsibility	Accountability
Privacy	Privacy	Privacy	Privacy and data governance
Autonomy	Freedom and autonomy	Freedom and autonomy	Human agency and oversight
Power over people	-	-	-
-	Trust	-	-
-	Sustainability	-	-
-	Dignity	-	-
-	Solidarity	-	-
-	-	Pedagogical appropriateness	-
-	-	Children's rights	-
-	-	AI literacy	-

As can be seen, the principles covered in the study at hand cover the core of the ethical principles identified in two literature reviews, and are fully consistent with the ethical guidelines of the European Commission. We complemented these principles with the principle of power.

2.1.4 Responsibility and accountability

Responsibility is about the obligation to satisfactorily perform a task; accountability is about liability for ensuring a task is satisfactorily done (McGrath & Whitty, 2018). This concerns both legal and moral aspects. In the realm of legal responsibility and accountability, the European Commission has developed a proposal for a regulatory framework on AI in the European Union (AI Act). In December 2023, the European Parliament and the Council reached agreement on the AI Act (European Commission, 2023). In addition to this legal side of responsibility and accountability, it needs to be clarified who should be responsible and accountable in the process of using AI. Attributing moral responsibility and accountability to individuals can be difficult, especially with AI, as it is often a black box (Long & Magerko, 2020). The question arises as to what extent taking responsibility and accountability is possible (for teachers) if no insight into the processes is attainable. The principle of accountability is part of the EGAIE.

2.1.5 Privacy

Privacy refers to the right of individuals to keep information about themselves secret (Manheim & Kaplan, 2019). There is a great incentive for organizations to collect comprehensive data. Data is the basis for AI applications (Wahlster, 2017). Users of AI services often have an incomplete understanding of the risks associated with disclosing private data. The trade-off between convenience and privacy is often in favor of convenience (Acquisti, Brandimarte & Loewenstein, 2015), which leads to people revealing data to organizations they reportedly distrust. Privacy is also at risk from hacking, or from exploiting security vulnerabilities (Lutz et al., 2019). AI can deeply invade the privacy of users. This is especially true for social robots that are equipped with extensive sensor technology. Security gaps and privacy threats are well documented in this area (Giaretta, Donno & Dragoni, 2018; Hoy, 2018). Privacy may be an important issue in education as sensitive data about learners is processed (European Commission, 2022).

2.1.6 Autonomy

The literature review of Jobin et al. (2019) identifies freedom and autonomy as an ethical principle. Moreover, the EGAIE covers the relationship between human beings and AI (European Commission, 2022). In this vein, the concept of autonomy, in particular, may be important to consider. Autonomy describes the ability of an agent to choose their own actions (Smakman et al., 2021). Autonomy in the context of AI can be classified based on the degree of automation (Molenaar, 2021): AI may only become active if a teacher prompts it, supporting the teacher or performing tasks under the supervision of the teacher. On the upper end of the autonomy continuum, AI fully performs tasks on its own. Different levels of autonomy can be distinguished,

ranging on a continuum from no autonomy to full autonomy (Guggemos et al., 2022a). The EGAIE opt against full autonomy although it might be possible from a technical point of view. Hence, in education fully autonomous AI may not be regarded as ethical; rather, the teacher may always be in the loop or in command (High-Level Expert Group on Artificial Intelligence, 2019).

2.1.7 Power over people

Power over people is not part of the literature review of Jobin et al. (2019). However, power is of key importance in social encounters; it can be defined as “potential influence” (Raven, 1993, p. 232). It is also a crucial concept in education (Niesche & Haase, 2012). Teachers can exert power over students, e. g., by their ability to reward them or through their expert knowledge. Depending on the level of autonomy, AI can exert power over teachers and learners as it can influence them. Power can occur in the form of hard and soft power (Busch & Link, 2021). A manifestation of soft power is expert knowledge, and of hard power the ability to sanction behavior (Raven, 1993). The key question concerning power may be whether or to what extent AI should be allowed to exert power over teachers and students.

2.2 Variables associated with teachers’ attitudes

Concerning technology integration in education, gender, age, and main subject are identified as potential sources of variance (Siddiq, Scherer & Tondeur, 2016; Šabić, Baranović & Rogošić, 2021). The reason for this may be differences in self-efficacy and, consequentially, attitudes. In terms of gender, males tend to report higher levels of technology self-efficacy. Concerning age, older teachers may have had less exposure to technology leading to lower self-efficacy in using technology. Regarding main subject, teachers may also have varying levels of exposure to technology. Subjects that inherently require a high level of technology integration, such as computer science, may lead to higher levels of technology self-efficacy among teachers compared to subjects that rely less on technology. Besides these context variables, the relationship between attitudes and knowledge is of concern (Scherer, Tondeur, Siddiq & Baran, 2018); however, the direction is unclear. On the one hand, knowledge about AI could lead to more critical attitudes towards AI due to an awareness of the challenges of such technology use, e. g., the issue of privacy. On the other hand, increasing knowledge may also lead to more favorable attitudes towards AI because people are becoming aware of their benefits. Moreover, positive attitudes may also lead to higher levels of knowledge, as teachers may be more open to learning about AI. The association between knowledge and attitudes may be bidirectional. Overall, there is broad evidence for a relationship between knowledge and attitudes (Evans & Durant, 1995); however, the sign and the direction of the association is unclear.

2.3 The present study

For the constructs presented in the previous section, the first author of the present study developed various items. These items were discussed and revised in the author team. Moreover, an external expert provided feedback. Table 2 presents the final set of items of the Artificial Intelligence Teachers' Ethical Attitudes scale (AI-TEA scale). We did not include items on non-maleficence because they overlap with other principles. Harm can evolve from neglecting other ethical principles, such as privacy and fairness. Including non-maleficence would impair the discriminant validity (AERA et al., 2014) of the measurement instrument, and a sound measurement instrument is the basis for quantitative empirical research.

Our first research question (RQ) is:

- RQ1: How can teachers' attitudes towards ethical principles be measured in a reliable and valid way?

In the case of ethical principles as a multifaceted concept (Jobin et al, 2019; European Commission, 2022), it is unlikely to observe only one profile. Rather, it can be expected to identify different sets of attitudes towards ethical principles among teachers in education. Knowledge about such profiles could be helpful in reducing complexity. Persons in the same profile could benefit from similar treatments (Hofmans, Wille & Schreurs, 2020); teacher education may develop instructional measures around the identified profiles.

Our second research question is:

- RQ2: Which latent profiles of pre-service teachers' attitudes towards ethical principles can be identified?

Furthermore, we are interested in what variables are associated with latent profile membership. This is an important step in predicting latent profile membership, which allows for a better understanding of the composition of these subgroups and how individuals are distributed among them. By identifying factors that can predict profile membership, it is possible to better understand heterogeneous experiences and therefore design more personalized support (Scherer et al., 2021). As reported in section 2.2, we expect gender, age, main subject, and AI knowledge to be correlated with latent profile membership:

- RQ3: How is latent profile membership related to contextual variables (gender, age, study subject) and AI knowledge?

Table 2: Items to operationalize the constructs.

Construct and item stem	Item (No.)
Autonomy	only take action when explicitly instructed to do so
Artificial intelligence applications should ...	by a teacher. (1)
	support teachers in their activities. (2)
	carry out activities under the supervision of the teacher. (3)
Justice and fairness	treat people fairly. (4)
Artificial intelligence applications should ...	give special support to disadvantaged people. (5)
	contribute to the elimination of existing inequalities. (6)
Soft power over teachers (formative)	make suggestions. (7)
Artificial intelligence applications should be allowed to help teachers ...	give formative feedback on their performance. (8)
Hard power over teachers (summative)	give binding instructions. (9)
Concerning teachers, artificial intelligence applications should be allowed to ...	give performance appraisals. (10)
	impose sanctions. (11)
Soft power over students (formative)	provide formative feedback on their performance. (12)
Concerning students, artificial intelligence applications should be allowed to ...	make suggestions. (13)
Hard power over students (summative)	give binding instructions. (14)
Concerning students, artificial intelligence applications should be allowed to ...	give grades. (15)
	impose sanctions. (16)
Privacy	access to data should be strictly controlled. (17)
When using artificial intelligence applications, ...	the dangers of hacker attacks should be minimised. (18)
Responsibility	the school supervisory authority. (19)
The responsibility for the negative consequences of the use of artificial intelligence applications should be borne by ...	the school management (20)
	the teacher. (21)
Transparency (process)	it should be disclosed that AI is being used. (22)
When using artificial intelligence applications ...	the data being used should be disclosed. (23)
	the source code should be disclosed. (24)
	the person/organisation responsible for its use should be disclosed. (25)
Transparency (data)	the developer company. (26)
Data generated during the use of artificial intelligence applications should be passed on in anonymized form to ...	the school management. (27)
	the school supervisory authorities. (28)
	the training company, if applicable. (29)
	the parents of students. (30)
	publicly accessible storage services. (31)
Well-being	promote their well-being. (32)
Concerning people, artificial intelligence applications should ...	support them in their concerns. (33)
	help with difficulties. (34)

3 Sample and Method

3.1 Data collection process

Given the conceptual basis that is closely aligned with the European Commission's ethical guidelines of AI in education, this study focuses on students who aspire to become teachers (prospective teachers). We collected data at eight universities in Germany and Switzerland in 2022 and 2023. The procedure of the data collection followed the ethical guidelines of our institutions; the participation in the study was entirely voluntary. We excluded all participants, 37 in number, who reported that they did not intend to become a teacher. Moreover, based on Mahalanobis distances and an inspection of the data, we excluded three more students. The final sample comprises 90 prospective teachers from vocational education and training at eight universities in Germany and Switzerland. The majority (58.9 %) are business and economics education students. The remainder are nursing and engineering education students. On average, the students are 29.03 years old ($SD = 5.28$ years), in their 2.93th semester ($SD = 1.76$) of their master's program, and have overall been studying for 9.86 semesters ($SD = 2.16$). Of the students, 53 % identify themselves as females.

3.2 Assessment of measurement model (RQ1)

Due to the rather small sample, we rely on the PLS-SEM approach to assess the quality of the measurement model. To this end, we use SmartPLS 4.0.9.3. A ρ greater than .7 points to a sufficient internal consistency reliability; an average variance extracted (AVE) greater than .5 might indicate convergent validity, and a heterotrait–monotrait ratio smaller than 0.85 discriminant validity (Sarstedt et al., 2022). After having identified the final model, we extract factor scores for the subsequent latent profile analysis (Scherer et al., 2021). Factor scores are superior to sum scores (McNeish & Wolf, 2020). We use unstandardized factor scores because they keep the original scale ranging from 0 (strongly disagree) to 5 (strongly agree). This facilitates the interpretation of the level of the constructs. To this end, we used the factor weighting scheme of the PLS algorithm (Chin, 1998).

3.3 Latent profile analysis (RQ2)

The approach of the latent profile analysis follows Guggemos, Seufert and Román-González (2023). We use the 'tidyLPA 1.1.0' R-package in combination with MPlus 8 to identify profiles of teachers' attitudes towards ethical aspects of AI use by means of a latent profile analysis (LPA) (Hallquist & Wiley, 2018; Rosenberg, Beymer, Anderson, van Lissa & Schmidt, 2018). We have to restrict variances to be equal across profiles and the covariance among the variables to be zero in order to avoid convergence problems (Meyer & Morin, 2016). The critical step in the LPA is to identify

an appropriate number of profiles. This decision might be based on information criteria and likelihood ratio tests, as well as on conceptual deliberations (Erhard, 2018; Scherer et al., 2021). Against this backdrop, we first assessed different class solutions. Following Morin and Marsh (2015) and Hofmans et al. (2020), we report the information criteria Akaike's information criterion (AIC), Consistent Akaike's information criterion (CAIC), Bayesian information criterion (BIC), and Adjusted Bayesian information criterion (aBIC), as well as the bootstrap likelihood ratio test (BLRT). Since these criteria may point to a different number of optimal profiles, we also utilize the approach of Akogul and Erisoglu (2017) where information criteria are weighted to determine the optimal number of latent profiles (from an empirical point of view). The identified solution should have a sufficiently high precision of classification, indicated by an entropy greater than 0.7. Besides this, the profiles should be substantially different from each other, which can be checked by means of a MANOVA (Tondeur et al., 2019). Afterwards, we evaluate if this approach yields a meaningful solution. The profiles should be of reasonable size and show substantial shape differences, i. e., differences that not only differ in levels but also in their pattern (Morin & Marsh, 2015).

3.4 Association with latent profile membership (RQ3)

To predict the profile membership, a multinomial regression would be suitable (Scherer et al., 2021); however, this is not possible due to our small sample. Hence, we rely on bivariate tests to evaluate the association between profile membership and context variables. For nominal scaled variables, e. g., gender, we use χ^2 -tests. In all other cases we rely on Kruskal-Wallis tests. To gauge the effect size of the associations, we use Cramer's V (χ^2 -tests) and η^2 (Kruskal-Wallis tests).¹ Reporting effect sizes is important; substantial effect sizes that are not statistically significant would suggest further research using a sample that yields higher statistical power in comparison to the study at hand (Fritz, Morris & Richler, 2012). On the other hand, statistically significant effects would not be practically meaningful if the effect size was small.

We measure AI knowledge using the test of Schmidt (2024). This test uses 21 single-choice items and focuses on basic knowledge of AI. In terms of content, the subject is subdivided into (1) basics of AI, (2) methods of AI, and (3) applications of AI. We use the sum score to assess AI knowledge. The internal consistency reliability is sufficiently high (Cronbachs $\alpha = .775$).

¹ For η^2 , a small effect may be 0.01, a medium effect 0.06, and a large effect 0.14 (Fritz et al., 2012). Concerning Cramer's V, a weak effect may be 0.05, a moderate effect 0.1, and a strong effect 0.15 (Akoglu, 2018).

4 Results

4.1 Assessment of measurement model

We had to make some adjustments in order to ensure a sound measurement model (see Figures 1 and 2).

Concerning the convergent validity of an overall power construct, we observe an AVE of .365. In line with the conceptual basis, the factor loadings indicate that power comprises both hard and soft types of power. First, AI can provide advice and (formative) feedback – an example of soft power. Second, AI can command or summatively assess teachers and students – hard power. After splitting up the constructs of teachers and students into hard and soft types of power, we observed a lack of discriminant validity for soft power (HTMT = 1.08). In other words, for the prospective teachers it does not make a difference whether AI exerts soft power over either teachers or students. However, also combining the hard power factor of teachers and students leads to an AVE of the new factor of .385. Hence, the data structure can be best described with a combined factor for soft power for teachers and students, and separate factors for teachers and students in the case of hard power. Figure 1 summarizes the development of the final model for power.

Concerning transparency, we had to split up the two covered facets: transparency about the process, e. g., informing about the use of AI; and transparency about the data, e. g., making the data available for research purposes. An AVE of .345 for the overall transparency construct indicates that transparency about the process and about the data are different constructs. Figure 2 depicts the adjustments concerning transparency.

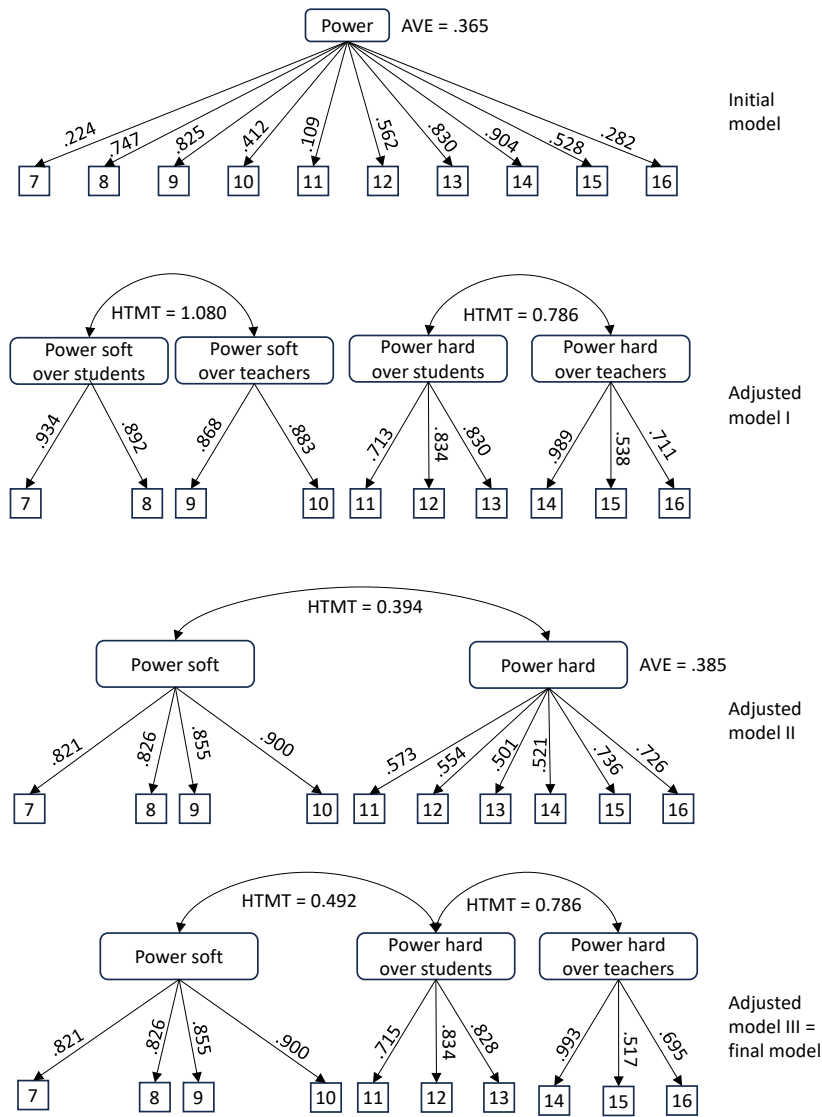


Figure 1: Adjustment from the initial model to the final model for power. The items (numbers 1 to 16) can be found in Table 2. The reported loadings are standardized. AVE = Average Variance Extracted; HTMT = heterotrait – monotrait ratio.

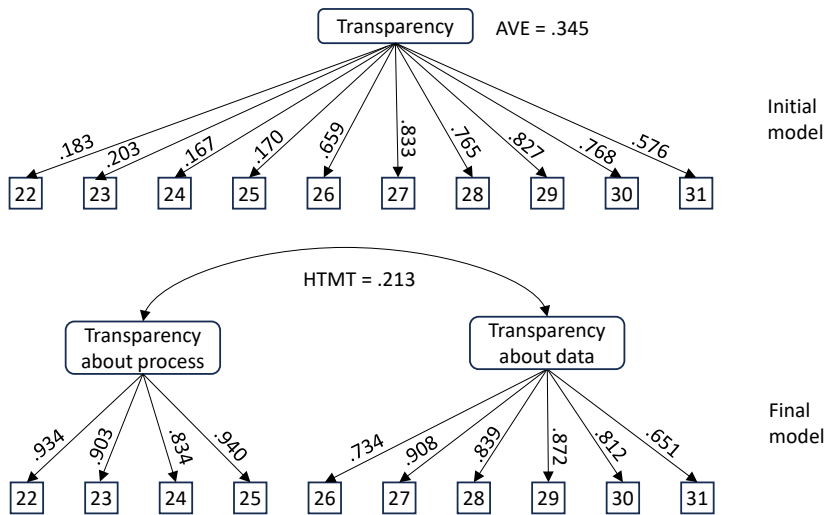


Figure 2: Adjustment from the initial model to the final model for transparency. The items (numbers 22 to 31) can be found in Table 2. The reported loadings are standardized. AVE = Average Variance Extracted; HTMT = heterotrait–monotrait ratio.

Table 3 reports the characteristics of the final models. Internal consistency reliability and convergent validity are sufficiently high ($\alpha > .7$, AVE $> .5$). Moreover, a HTMT smaller than .553 ($< .85$) indicates discriminant validity.

4.2 Latent profile analysis

Table 4 shows the results of the latent profile analysis; the information criteria point to different class solutions. The algorithm of Akogul and Erisoglu (2017) and the bootstrap likelihood ratio test suggests three profiles as the optimum. Moreover, the entropy for this solution is sufficiently high (0.909). The MANOVA indicated significant different means among the three profiles: $F(16, 160) = 50.16$, Wilk's $\Lambda = 0.115$, $p < .001$, $\eta^2 = .588$. In other words, 58.8 % of the variance can be explained by profile membership. These findings and Figure 3 point to sufficiently distinct profiles. Solutions with more than three profiles yield very small samples with only two prospective teachers.

Table 3: Assessment of measurement model and correlations among constructs.

Variable	M	SD	pa	AVE	1	2	3	4	5	6	7	8	9
1. autonomy	3.30	0.75	.77	.64									
2. justice & fairness	3.98	0.83	.87	.79	.23								
3. well-being	3.60	0.71	.87	.78	.14	.24							
4. power (soft)	3.33	0.86	.90	.72	.23	.40	.38						
5. power teachers (hard)	1.49	0.91	.78	.58	-.12	-.18	.10	.09					
6. power students (hard)	1.74	1.08	.71	.63	-.11	-.12	.10	.31	.66				
7. privacy	4.10	0.79	.76	.74	-.52	-.03	.11	-.54	-.25	-.52			
8. responsibility	2.20	0.70	.80	.66	.24	.10	.23	.08	.18	.06	.07		
9. transparency (process)	3.64	1.04	.93	.82	.55	.70	.25	.09	-.63	-.59	.20	.41	
10. transparency (data)	1.91	0.81	.93	.65	.09	-.03	.30	.23	.26	.21	-.14	.03	-.92

Note: Constructs measured on a six-point scale of rating ranging from 0 (strongly disagree) to 5 (strongly agree). Correlations greater than |.22| are statistically significant at the 5 % level (two-sided). M = Mean, SD = Standard Deviation, pa = measure for internal consistency reliability, AVE = Average Variance Extracted.

Table 4: Information criteria, entropies, and BLRT results for one to five latent profiles.

No. profiles	-LL	No. Par.	AIC	CAIC	BIC	aBIC	Entropy	p(BLRT)
1	938	20	1917	1987	1967	1904	1	-
2	902	31	1865	1974	1943	1845	0.876	.000
3	874	42	1831	1978	1936	1804	0.909	.000
4	857	53	1819	2005	1952	1785	0.932	.154
5	845	64	1819	2043	1979	1777	0.899	.600

Notes: -LL = -Log-likelihood, No. Par. = number of estimated parameters, AIC = Akaike's information criterion, CAIC = Consistent AIC, BIC = Bayesian Information Criterion, aBIC = adjusted BIC, BLRT = bootstrap likelihood ratio test.

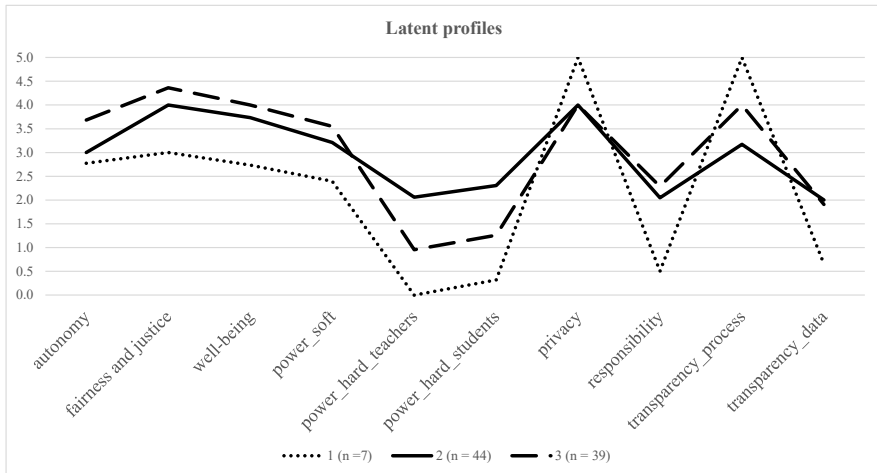


Figure 3: Description of the three latent profiles (N = 90).

The three latent profiles can be described as follows:

- ▶ Profile 1: Prospective teachers in this profile have high standards for AI use in terms of privacy and transparency of the process. Moreover, they entirely disagree that AI should exert power over teachers and students. They also disagree that the school (both teachers and principals) should take over responsibility for potential (negative) outcomes of AI use. Profile 1 is the smallest one with 7 prospective teachers (7.8 %).
- ▶ Profile 2: This profile is characterized by its neutral to slightly negative attitude towards the use of hard power of AI over teachers and students. Profile 2 is the largest one with 44 prospective teachers (48.9 %).
- ▶ Profile 3: This profile is in general similar to profile 2. The difference between the two profiles is the more pronounced disagreement that AI should exert hard power over teachers and students. Profile 3 comprises 39 prospective teachers (43.3 %).

4.3 Association with latent profile membership

There is no significant relationship of profile membership with gender ($V = .073$, $p = .780$), age ($\eta^2 = 0.01$, $p = .600$), study subject of business and economics education vs. all other study subjects ($V = .078$, $p = .758$), and AI knowledge ($\eta^2 = 0.00$, $p = .417$). Figure 4 depicts the results of the knowledge test grouped by the three latent profiles. The 95 % confidence interval of profile 1, on the one hand, and profiles 2 and 3, on the other, do not overlap.

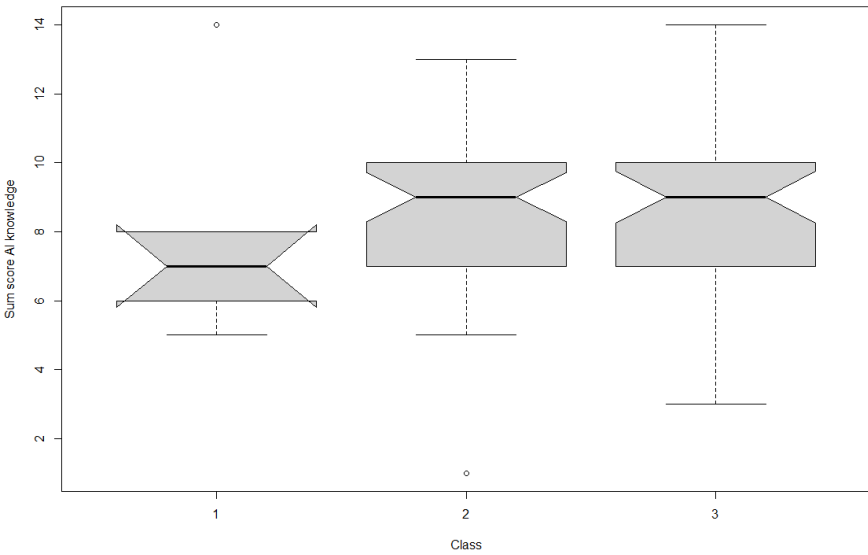


Figure 4: AI knowledge grouped by latent profile membership (notches indicate 95 % confidence intervals).

5 Discussion and Limitations

We have developed an instrument to assess (prospective) teachers' attitudes towards the ethical principles of AI use in education (RQ1), identified latent profiles of prospective teachers towards these principles (RQ2), and investigated the association of latent profile membership with context variables and AI knowledge (RQ3). Concerning RQ1, we had to adjust the initial instruments in order to achieve a sufficient internal consistency reliability, as well as discriminant and convergent validity. These adjustments can be justified from a conceptual point of view. Power can be classified as soft and hard power (Busch & Link, 2021). Regarding soft power, the prospective teachers do not differentiate between whether teachers or students are concerned; however, the same teachers do see a difference in whether hard power is exerted over teachers or over students ($d = 0.31$, $p < .01$). From an ethical point of view this is interesting. The golden rule can be stated as: "Treat others only as you consent to being treated in the same situation" (Gensler, 2013, p. 2). The question arises: if the prospective teachers stepped into their students' shoes, would they want AI to exert power over them? Against this backdrop, it would be promising to investigate a student sample in order to see if they would be willing to allow AI to exert power over both them and their teachers. In terms of transparency, the pro-

spective teachers value privacy about the process, 3.64 ($SD = 1.04$), but are skeptical towards transparency about the data, 1.19 ($SD = 0.81$). Transparency about the data and transparency about the process are highly negatively correlated, $r = -.92$ ($p < .001$). Although the constructs can be measured with a sufficient discriminant validity, it may be suitable to remove one of these transparency scales from the instrument. Privacy about the process is measured with only two items. Although this yields a sufficient internal consistency reliability ($\rho_a = .76$), it would be beneficial to include further items. This also applies to autonomy, as well as to hard power over students and teachers ($.7 < \rho_a < .8$).

A general limitation is our reliance on self-reports to capture prospective teachers' ethical beliefs. The validity of self-reports can be called into question (Wojcik, Hovasapian, Graham, Motyl & Ditto, 2015; SWK, 2022). An alternative could be to use web scraping (Wojcik et al., 2015) to analyze the tweets of teachers on X (formerly Twitter) that cover AI, in order to investigate their ethical beliefs. As a qualitative method, the critical incident technique may be a viable option to make inferences about teachers' ethical beliefs. Despite these limitations, our instrument allows a reliable assessment of key ethical aspects when using AI in education. Researchers and teacher trainers can use it to capture (prospective) teachers' attitudes towards the ethical principles of artificial intelligence use in education.

Concerning RQ2, prospective teachers in all profiles agree concerning the values of fairness and justice, well-being, privacy, and transparency about the process. In terms of taking over responsibility and accountability as well as transparency about the data, the prospective teachers' attitudes can be rated overall as negative. The main difference between the profiles is whether prospective teachers agree that AI should exert hard power over teachers and students. In particular, the prospective teachers in profile 1 are very skeptical about AI exerting hard power over people. Although profile 1 is rather small (7.8 %), it may be important to consider these concerns when integrating AI into education. Interestingly, prospective teachers in profile 1 have smaller values for justice and fairness, as well as well-being. This might be because they have a critical stance towards AI and think such technology is not able to ensure justice, fairness, and well-being. The critical stance towards AI of the prospective teachers in profile 1 may also be reflected in high values for privacy and transparency for the process. The size of this profile is consistent with the findings of Guggemos, Sonderegger, and Seufert (2022b), who reported a profile of students of similar size (12.5 %) who are very skeptical towards the use of social robots (manifestation of AI) in education.

Concerning RQ3, we did not find a statistically significant association of latent profile membership with the context variables. Moreover, the effect sizes are small to negligible. Hence, using a sample that yields higher statistical power than our sample

($N = 90$) may also not reveal effects of practical relevance. The relationship between AI knowledge and latent profile membership is also not statistically significant. However, the confidence intervals of profile 1, on the one hand, and profiles 2 and 3, on the other, do not overlap. People who lack knowledge about a subject may be skeptical towards it. However, there is also an outlier in profile 1 (see Figure 2): a prospective teacher who achieved the highest score in the AI knowledge test (14 out of 21). A hypothesis would be that the association between AI knowledge and profile membership follows an inverted U. With increasing knowledge, students become less skeptical towards AI. However, when the level of knowledge reaches a specific point, the positive attitudes decrease. AI experts, in particular, issue warnings about the risks of AI (Center for AI Safety, 2023). This hypothesis is in line with the findings of Zeng, Wei, Zhu, Zhao and Lin (2019).

Due to its small sample size, our study is exploratory in nature. Furthermore, we only sampled prospective teachers in vocational education and training, mainly students in business and economics. It would also be interesting to use the developed instrument with (prospective) teachers from other fields. We addressed AI in general; however, attitudes towards the ethical principles of AI may depend on the kind of technology used, e. g., a social robot vs. an intelligent tutorial system. Latent profile analysis is an explorative technique. In combination with our small sample with a limited scope, it would be beneficial to replicate our results.

6 Conclusion and Outlook

This paper sheds light on the attitudes of prospective teachers towards the ethical principles of the use of AI in education. Attitudes are an important driver of technology acceptance. Based on a newly developed instrument (AI-TEA scales), we assessed the constructs of transparency, justice and fairness, well-being, responsibility, privacy, autonomy, and power over people. The findings contribute to the discourse on the responsible and ethical use of AI in education, in particular within the European Union. It provides a foundation for future research in this area. In particular, the findings indicate that AI that exerts hard power over teachers may face severe challenges of technology acceptance among teachers, i. e., they may refuse to use it. However, teachers may accept formative feedback from AI: prospective teachers from all three profiles agree that AI should be allowed to provide feedback. Our study investigates the association of AI knowledge and teachers' attitudes towards ethical aspects in education. This AI knowledge is generic in nature. Future research should focus on specific knowledge concerning ethical aspects, for instance, student knowledge about the bias and unfairness of AI. Overall, the relationship between AI knowledge and attitudes seems to be a promising avenue for further research. A further interesting point is to discuss whether knowledge about the ethical issues of

AI should be part of AI literacy (Laupichler, Aster & Raupach, 2023; Long & Magerko, 2020; Ng, Leung, Chu & Qiao, 2021). It would also be promising to investigate whether available AI applications, such as intelligent tutoring systems and social robots, would vary in their acceptance by teachers. Besides (prospective) teachers, it may be insightful to also explore the attitudes towards the ethical principles of further important stakeholders, such as parents and school principals.

References

- Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, 347 (6221), 509–514. <https://www.doi.org/10.1126/science.aaa1465>
- Adams, C., Pente, P., Lernermeier, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence*, 4, 100131. <https://www.doi.org/10.1016/j.caeai.2023.100131>
- AERA, APA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association. Available under: <https://eric.ed.gov/?id=ED565876> [08.03.2024]
- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18 (3), 91–93. <https://www.doi.org/10.1016/j.tjem.2018.08.001>
- Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2 (3), 431–440. <https://www.doi.org/10.1007/s43681-021-00096-7>
- Akogul, S., & Erisoglu, M. (2017). An approach for determining the number of clusters in a model-based cluster analysis. *Entropy*, 19 (9), 452. <https://www.doi.org/10.3390/e19090452>
- André, E., Aurich, J. C., Bauer, W., Bullinger-Hoffmann, A., Heister, M., Huchler, N., Neuburger, R., Peissner, M., Stich, A., Suchy, O., Ramin, P., & Wächter, M. (2021). Kompetenzentwicklung für KI: Veränderungen, Bedarfe und Handlungsoptionen. Plattform Lernende Systeme. Whitepaper. https://www.doi.org/10.48669/pls_2021-2
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, 3 (21), 1–9. <https://www.doi.org/10.1126/scirobotics.aat5954>
- Busch, M. W., & Link, K. (2021). Agile Organisationen – Versuch einer kritischen Bestandsaufnahme. *Journal Für Psychologie*, 29 (1), 9–38. <https://www.doi.org/10.30820/0942-2285-2021-1-9>
- Center for AI Safety (2023). Statement on AI Risk: AI experts and public figures express their concern about AI risk. Available under: <https://www.safe.ai/work/statement-on-ai-risk> [08.03.2024].
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264–75278. <https://www.doi.org/10.1109/ACCESS.2020.2988510>
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern Methods for Business Research* (pp. 295–336). Mahwah: Lawrence Erlbaum.
- Erhard, F. (2018). Estimating pseudocounts and fold changes for digital expression measurements. *Bioinformatics*, 34 (23), 4054–4063. <https://www.doi.org/10.1093/bioinformatics/bty471>
- European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Brussels: Publications Office of the European. <https://www.doi.org/10.2766/153756>
- European Commission. (2023). Commission welcomes political agreement on Artificial Intelligence Act. Available under: https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473 [08.03.2024].
- Evans, G., & Durant, J. (1995). The relationship between knowledge and attitudes in the public understanding of science in Britain. *Public Understanding of Science*, 4 (1), 57–74. <https://www.doi.org/10.1088/0963-6625/4/1/004>
- Felzmann, H., Fosch-Villaronga, E., Lutz, C. & Tamò-Larrieux, A. (2020). Towards Transparency by Design for Artificial Intelligence. *Science and Engineering Ethics*, 26 (6), 3333–3361. <https://www.doi.org/10.1007/s11948-020-00276-4>

- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P. & Vayena, E. (2018). Ai4people-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds & Machines*, 28 (4), 689-707. <https://www.doi.org/10.1007/s11023-018-9482-5>
- Fritz, C. O., Morris, P. E. & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology*, 141 (1), 2-18. <https://www.doi.org/10.1037/a0024338>
- Gensler, H. J. (2013). *Ethics and the golden rule*. New York: Routledge.
- Giaretta, A., Donno, M. D. & Dragoni, N. (2018). Adding Salt to Pepper: A Structured Security Assessment over a Humanoid Robot. *Proceedings of International Conference on Availability, Reliability and Security*. (pp. 1-8). <https://www.doi.org/10.1145/3230833.3232807>
- Guggemos, J., Seufert, S. & Román-González, M. (2023). Computational Thinking Assessment – Towards More Vivid Interpretations. *Technology, Knowledge and Learning*, 28 (2), 539-568. <https://www.doi.org/10.1007/s10758-021-09587-2>
- Guggemos, J., Seufert, S., Sonderegger, S. & Burkhard, M. (2022a). Social Robots in Education: Conceptual Overview and Case Study of Use. In D. Ifenthaler, P. Isaías & D. G. Sampson (Eds.), *Orchestration of Learning Environments in the Digital World* (pp. 173-195). Cham: Springer. https://www.doi.org/10.1007/978-3-030-90944-4_10
- Guggemos, J., Sonderegger, S., & Seufert, S. (2022b). Student Acceptance of Social Robots in Higher Education: Evidence from a Vignette Study. In D. Sampson, D. Ifenthaler, & P. Isaías (Eds.), *19th International Conference on Cognition and Expository Learning in Digital Age (CELDA)* (pp. 105-114). Lisbon: IADIS Press.
- Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Waldron, L., Wang, B., McIntosh, C., Goldenberg, A., Kundaje, A., Greene, C. S., Broderick, T., Hoffman, M. M., Leek, J. T., Korthauer, K., Huber, W., Brazma, A., Pineau, J., Tibshirani, R., Hastie, T., . . . Aerts, H. J. W. L. (2020). Transparency and reproducibility in artificial intelligence. *Nature*, 586 (7829), E14-E16. <https://www.doi.org/10.1038/s41586-020-2766-y>
- Hallquist, M. N. & Wiley, J. F. (2018). *Mplusautomation: An R package for facilitating large-scale latent variable analyses in Mplus*. *Structural Equation Modeling: A Multidisciplinary Journal*, 25 (4), 621-638. <https://www.doi.org/10.1080/10705511.2017.1402334>
- Harel, A., Shabtai, A., Rokach, L., & Elovici, Y. (2010). M-Score: Estimating the Potential Damage of Data Leakage Incident by Assigning Misuseability Weight. *Proceedings of the 9th Annual International ACM Conference on International Computing Education Research*, 13-20. <https://www.doi.org/10.1145/1866886>
- Heller, M. A. & Phillips, S. A. (2020). A new ethical beliefs scale. *Ethics & Behavior*, 30 (7), 496-513. <https://www.doi.org/10.1080/10508422.2019.1668792>
- High-Level Expert Group on Artificial Intelligence. (2019). *Ethics I for trustworthy AI*. European Commission.
- Hofmans, J., Wille, B. & Schreurs, B. (2020). Person-centered methods in vocational research. *Journal of Vocational Behavior*, 118, 103398. <https://www.doi.org/10.1016/j.jvb.2020.103398>
- Holmes, W., Bialik, M. & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I. & Koedinger, K. R. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32 (3), 504-526. <https://www.doi.org/10.1007/s40593-021-00239-1>
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37 (1), 81-88. <https://www.doi.org/10.1080/02763869.2018.1404391>
- Hwang, G.-J., Xie, H., Wah, B. W. & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1, 100001. <https://www.doi.org/10.1016/j.caeai.2020.100001>

- Instefjord, E. J., & Munthe, E. (2017). Educating digitally competent teachers: A study of integration of professional digital competence in teacher education. *Teaching and Teacher Education*, 67, 37-45. <https://www.doi.org/10.1016/j.tate.2017.05.016>
- Jobin, A., Ienca, M. & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1 (9), 389-399. <https://www.doi.org/10.1038/s42256-019-0088-2>
- Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103 (3), 582-591. <https://www.doi.org/10.1037/0033-295x.103.3.582>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Knezek, G. & Christensen, R. (2016). Extending the will, skill, tool model of technology integration: adding pedagogy as a new model construct. *Journal of Computing in Higher Education*, 28 (3), 307-325. <https://www.doi.org/10.1007/s12528-016-9120-2>
- Laupichler, M. C., Aster, A. & Raupach, T. (2023). Delphi study for the development and preliminary validation of an item set for the assessment of non-experts' AI literacy. *Computers and Education: Artificial Intelligence*, 4, 100126. <https://www.doi.org/10.1016/j.caeai.2023.100126>
- Lindner, A., & Berges, M.-P. (2020). Can you explain AI to me? Teachers' pre-concepts about Artificial Intelligence. 2020 IEEE Frontiers in Education Conference (FIE).
- Long, D. & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. In R. Bernhaupt (Ed.), *ACM Digital Library, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). Association for Computing Machinery. <https://www.doi.org/10.1145/3313831.3376727>
- Lutz, C., Schöttler, M. & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media & Communication*, 7 (3), 412-434. <https://www.doi.org/10.1177/2050157919843961>
- Manheim, K. & Kaplan, L. (2019). Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law and Technology*, 21, 106-188.
- Marcinkowski, F., Kieslich, K., Starke, C. & Lünich, M. (2020). Implications of AI (un-)fairness in higher education admissions. In M. Hildebrandt, C. Castillo, E. Celis, S. Ruggieri, L. Taylor & G. Zanfir-Fortuna (Eds.), *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency: January 27-30, 2020, Barcelona, Spain* (pp. 122-130). The Association for Computing Machinery. <https://www.doi.org/10.1145/3351095.3372867>
- McGrath, S. K. & Whitty, S. J. (2018). Accountability and responsibility defined. *International Journal of Managing Projects in Business*, 11 (3), 687-707. <https://www.doi.org/10.1108/IJMPB-06-2017-0058>
- McNeish, D. & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52 (6), 2287-2305. <https://www.doi.org/10.3758/s13428-020-01398-0>
- Meyer, J. P. & Morin, A. J. (2016). A person-centered approach to commitment research: Theory, research, and methodology. *Journal of Organizational Behavior*, 37 (4), 584-612. <https://www.doi.org/10.1002/job.2085>
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3 (2), 1-21. <https://www.doi.org/10.1177/2053951716679679>
- Molenaar, I. (2021). Personalisation of learning: Towards hybrid human-AI learning technologies. In OECD (Ed.), *OECD Digital Education Outlook 2021: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, OECD Publishing (pp. 57-77). Brussels: OECD.
- Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, 22 (1), 39-59. <https://www.doi.org/10.1080/10705511.2014.919825>

- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W. & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://www.doi.org/10.1016/j.caeai.2021.100041>
- Niesche, R. & Haase, M. (2012). Emotions and Ethics: A Foucauldian framework for becoming an ethical educator. *Educational Philosophy and Theory*, 44 (3), 276-288. <https://www.doi.org/10.1111/j.1469-5812.2010.00655.x>
- O'Donoghue, K. (2023). Learning Analytics within Higher Education: Autonomy, Beneficence and Non-maleficence. *Journal of Academic Ethics*, 21 (1), 125-137. <https://www.doi.org/10.1007/s10805-021-09444-y>
- Pargmann, J. & Berding, F. (2022). Künstliche Intelligenz in der beruflichen Bildung – Utopie oder Dystopie? Chancen und Herausforderungen von KI als Bildungstechnologie. In M. Eckelt, T. J. Ketschau, J. Klassen, J. Schauer, J. K. Schmees & C. Steib (Eds.), *Berufsbildungspolitik: Strukturen – Krise – Perspektiven* (pp. 215–232). Bielefeld: Wbv.
- Peissner, M., Kötter, F. & Zaiser, H. (2019). Künstliche Intelligenz – Anwendungsperspektiven für Arbeit und Qualifizierung. *BWP*, 48 (3), 9–13.
- Pega (2018). What Consumers Really Think About AI: A Global Study. Available under: <https://www.pegacom/ai-survey> [08.03.2024].
- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334 (6060), 1226-1227. <https://www.doi.org/10.1126/science.1213847>
- Raven, B. H. (1993). The Bases of Power: Origins and Recent Developments. *Journal of Social Issues*, 49 (4), 227-251. <https://www.doi.org/10.1111/j.1540-4560.1993.tb01191.x>
- Rawls, J. (1999). *A theory of justice* (Rev. ed.). Cambridge Mass.: Belknap Press of Harvard University Press.
- Rosenberg, J., Beymer, P., Anderson, D., van Lissa, C. & Schmidt, J. (2018). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, 3 (30), 978. <https://www.doi.org/10.21105/joss.00978>
- Šabić, J., Baranović, B. & Rogošić, S. (2021). Teachers' Self-efficacy for Using Information and Communication Technology: The Interaction Effect of Gender and Age. *Informatics in Education*, 21 (2), 353-373. <https://www.doi.org/10.15388/infedu.2022.11>
- Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L. & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing*, 39 (5), 1035-1064. <https://www.doi.org/10.1002/mar.21640>
- Scherer, R., Howard, S. K., Tondeur, J. & Siddiq, F. (2021). Profiling teachers' readiness for online teaching and learning in higher education: Who's ready? *Computers in Human Behavior*, 118, 106675. <https://www.doi.org/10.1016/j.chb.2020.106675>
- Scherer, R., Siddiq, F. & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. *Computers & Education*, 128, 13-35. <https://www.doi.org/10.1016/j.compedu.2018.09.009>
- Scherer, R., Tondeur, J., Siddiq, F. & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior*, 80, 67-80. <https://www.doi.org/10.1016/j.chb.2017.11.003>
- Schmidt, J. (2024). *Grundlagenwissen zu Künstlicher Intelligenz (angehender) Lehrkräfte: Modellbasierte Testentwicklung und Validierung*. Bielefeld: Wbv.
- Seufert, S., Guggemos, J. & Ifenthaler, D. (2021a). Zukunft der Arbeit mit intelligenten Maschinen: Implikationen der Künstlichen Intelligenz für die Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, (31. Beiheft), 9–27.
- Seufert, S., Guggemos, J. & Sailer, M. (2021b). Technology-related knowledge, skills, and attitudes of pre-and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 106552. <https://www.doi.org/10.1016/j.chb.2020.106552>

- Siddiq, F., Scherer, R. & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): A new construct in 21st century education. *Computers & Education*, 92-93, 1-14. <https://www.doi.org/10.1016/j.compedu.2015.10.006>
- Smakman, M., Vogt, P. & Konijn, E. A. (2021). Moral considerations on social robots in education: A multi-stakeholder perspective. *Computers & Education*, 174, 104317. <https://www.doi.org/10.1016/j.compedu.2021.104317>
- Ständige Wissenschaftliche Kommission (SWK) (2022). Digitalisierung im Bildungssystem: Handlungsempfehlungen von der Kita bis zur Hochschule: Gutachten der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz (SWK). <https://doi.org/10.25656/01:25273>
- Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, 20 (3), 571-610. <https://www.doi.org/10.5465/amr.1995.9508080331>
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68 (1), 1-16. <https://www.doi.org/10.1007/s11423-019-09701-3>
- Tambe, P., Cappelli, P. & Yakubovich, V. (2019). Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *California Management Review*, 61 (4), 15-42. <https://www.doi.org/10.1177/0008125619867910>
- Tondeur, J., Scherer, R., Baran, E., Siddiq, F., Valtonen, T. & Sointu, E. (2019). Teacher educators as gatekeepers: Preparing the next generation of teachers for technology integration in education. *British Journal of Educational Technology*, 50 (3), 1189-1209. <https://www.doi.org/10.1111/bjet.12748>
- Wahlster, W. (2017). Artificial intelligence as the driver for the second wave of digitalization. *IM+io*(2), 10-13. Available under: <https://www.vdmimpulse.org/article/-/article/render/156223> [20.05.2019].
- Wang, W. & Siau, K. (2019). Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity. *Journal of Database Management*, 30 (1), 61-79. <https://www.doi.org/10.4018/JDM.2019010104>
- Wojcik, S. P., Hovasapian, A., Graham, J., Motyl, M. & Ditto, P. H. (2015). Conservatives report, but liberals display, greater happiness. *Science*, 347 (6227), 1243-1246. <https://www.doi.org/10.1126/science.1260817>
- Zawacki-Richter, O., Marín, V. I., Bond, M. & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16 (1), 16-39. <https://www.doi.org/10.1186/s41239-019-0171-0>
- Zeng, J., Wei, J., Zhu, W., Zhao, D. & Lin, X. (2019). Residents' behavioural intentions to resist the nuclear power plants in the vicinity: an application of the protective action decision model. *Journal of Risk Research*, 22 (3), 382-400. <https://www.doi.org/10.1080/13669877.2017.1391316>

Autor*innen

Jun.-Prof. Dr. Josef Guggemos, PH Schwäbisch Gmünd, Juniorprofessur für Digitalisierung in der beruflichen Bildung

Jun.-Prof. Dr. Jacqueline Schmidt, Technische Universität Dresden, Juniorprofessur für Wirtschaftspädagogik, insb. Digitalisierung in Bildungs- und Arbeitswelten

Prof. Dr. Roland Happ, Universität Leipzig, Institut für Wirtschaftspädagogik, Professur für Berufliche Bildung mit dem Schwerpunkt Wirtschaft

Korrespondenz an: josef.guggemos@ph-gmuend.de

Marc Egloffstein, Kristina Kögler & Dirk Ifenthaler

Evidenzgestützte Entwicklung von onlinebasierten Lernangeboten zu Künstlicher Intelligenz in der beruflichen Bildung: Stakeholder-Perspektiven und Implementierung

Dieser Beitrag beschreibt die evidenzgestützte Entwicklung von onlinebasierten Lernangeboten zum Thema Künstliche Intelligenz (KI) in der beruflichen Bildung. Im Rahmen der Zielgruppen- und Kontextanalyse des Projekts AI_VET wurden leitfadengestützte Interviews mit N = 48 Stakeholder*innen aus der beruflichen Bildung durchgeführt und ausgewertet. Die Ergebnisse zeigen ein diffuses und eher oberflächliches Verständnis von KI innerhalb der Zielgruppe sowie eine mangelnde Verankerung von KI in der beruflichen Praxis, trotz eines breiten Konsenses hinsichtlich ihrer zunehmenden Relevanz. Konkrete Impulse für die Gestaltung von onlinebasierten Lernangeboten sind die Ausrichtung auf KI-Grundlagen, die Auseinandersetzung mit Chancen, Risiken und Limitationen in beruflichen Kontexten sowie die Nutzung konkreter Anwendungsbeispiele. Auf dieser Basis wurde die Kursreihe AI_VET in einem partizipativen Ansatz implementiert. Erste Evaluationsergebnisse deuten auf eine differenzierte Nutzung der Kursbausteine sowie auf Akzeptanz von Seiten der Lernenden hin. Implikationen für die weitere Verankerung von KI in der beruflichen Bildung sowie die Einordnung der Angebote in den Kontext der KI-Bildung werden diskutiert.

Schlagwörter: Berufliche Bildung – Instruktionsdesign – KI-Campus – Künstliche Intelligenz – Online-Lernen

Evidence-based development of online learning resources on Artificial Intelligence in vocational education and training: Stakeholder perspectives and implementation

This article describes the evidence-based development of online learning resources on Artificial Intelligence (AI) in vocational education and training. Within the target group and context analysis of the AI_VET project, guided interviews with N = 48 stakeholders from vocational education and training were conducted and analyzed. The results reveal an ambiguous and rather superficial understanding of AI within the target group as well as a limited incorporation of AI into vocational practice despite a broad consensus regarding its increasing relevance. Recommendations for the design of online learning resources include a focus on AI basics to promote a foundational understanding, consideration of opportunities, risks and limitations in professional contexts, and the use of specific application examples. Building on this, the AI_VET course series was implemented through a participatory approach. Initial evaluation results suggest differentiated usage of the course modules and acceptance among learners. The article concludes with a discussion of implications for the further integration of AI in vocational education and training, as well as the positioning of the developed offerings in AI education.

Keywords: AI Campus – Artificial Intelligence – instructional design – online learning – VET

1 Ausgangslage und Zielsetzung des Beitrags

Vom breiten Einsatz von Verfahren und Methoden der Künstlichen Intelligenz (KI) ist der kommende Innovationsschub innerhalb der digitalen Transformation von Wirtschaft und Gesellschaft zu erwarten (Wittpahl, 2019). Spätestens mit der sprunghaften Verbreitung von generativer KI seit Ende 2022 ist Künstliche Intelligenz auch zum Gegenstand von umfangreichen Erwägungen zur Gestaltung und Unterstützung von Bildungsprozessen geworden (Chiu, 2023). Die Europäische Kommission wies bereits 2013 auf die Potenziale hin, die aus dem Einsatz von KI-Technologien und Learning Analytics im Bildungsbereich erwachsen können. Damit verbunden war der Appell an die Institutionen der schulischen, beruflichen, hochschulischen und Erwachsenenbildung, die Nutzung von KI-gestützten Ansätzen in der Lehre konzeptionell voranzutreiben und zu unterstützen. Angesichts der Tragweite der Veränderungen in der Berufs- und Arbeitswelt durch KI und der aktuellen und zukünftigen Relevanz für berufliche Lehr-Lernprozesse liegt es nahe, das Thema mit seinen Potenzialen und Herausforderungen stärker in die berufs- und wirtschaftspädagogische Aus- und Weiterbildung zu integrieren (Seufert, Guggemos & Ifenthaler, 2021; Wilbers, 2020).

Für berufs- und wirtschaftspädagogische Ausbildungsgänge ist KI sowohl in inhaltlicher als auch in methodischer Hinsicht relevant. So werden Auswirkungen von KI auf unterschiedliche Berufsfelder diskutiert (Ertl & Seifried, 2021) und KI-basierte Anwendungsszenarien unter Labels wie ‚Industrie 4.0‘ oder ‚Smart Services‘ besprochen (Spöttl & Windelband, 2019). Die Implementierung KI-gestützter Bildungstechnologien war trotz aller Potenziale (Söllner, Janson, Rietsche & Thiel de Gafenco, 2021) bis in die jüngere Zeit hinein – mit vereinzelt Ausnahmen wie etwa Learning Analytics (Ifenthaler, 2020) – im Bereich der beruflichen Bildung vergleichsweise unterrepräsentiert (Attwell et al., 2020). Aus- und Weiterbildungsangebote für KI waren bis etwa 2020 nur sehr vereinzelt zu finden (Seufert et al., 2021). Seither ist allerdings eine dynamische Entwicklung zu verzeichnen, die sich in verschiedenen, auch bundesweiten Initiativen in Forschung und Lehre niederschlägt. In der beruflichen Bildung werden vermehrt Kooperationsprojekte initiiert, die darauf abzielen, Kompetenzen im Zusammenhang mit KI zu fördern.

Für eine zielführende Implementierung von KI in berufliche Bildungsprozesse müssen die Akteur*innen der beruflichen Bildung zum einen in der Lage sein, über KI als Bildungsgegenstand und Bestandteil des Bildungsplans zu lehren, was neben den KI-induzierten Veränderungen von Lehr-, Lern- und Arbeitsprozessen nicht zuletzt auch eine Auseinandersetzung mit Herausforderungen und Limitationen erfordert. Zum anderen müssen sie befähigt werden, KI-Anwendungen als methodische Werkzeuge zur Unterstützung des Lehrens und Lernens zielführend einzusetzen (Seufert, 2020). Vor dem Hintergrund der sich daraus abzeichnenden umfangreichen Aus-

und Weiterbildungsbedarfe werden in jüngerer Zeit vermehrt digitale Lernangebote zu KI entwickelt und ausgebaut (Mah et al., 2020; Ng et al., 2023). Hierbei spielen Massive Open Online Courses (MOOCs) als onlinebasierte Angebote für große Adressatengruppen sowohl in der Aus- als auch in der Weiterbildung eine übergeordnete Rolle (vgl. Ebner, Schön & Braun, 2019). MOOCs fungieren nicht nur als alleinstehende Bildungsangebote im Rahmen individueller (oder auch institutionalisierter) Weiterbildungsaktivitäten, sondern werden auch im Rahmen der hochschulischen Ausbildung von zukünftigem beruflichen Bildungspersonal eingesetzt. Hierbei zeichnen sich nicht nur sehr unterschiedliche Implementierungsmöglichkeiten etwa im Rahmen von Blended Learning-Szenarios (Littenberg-Tobias & Reich, 2020) oder modularen Ausbildungskonzepten (Serth, Staubitz, von Elten & Meinel, 2022) ab, sondern auch spezifische Herausforderungen, die mit der Heterogenität der Adressatengruppen bezüglich Vorkenntnissen, Teilnahmemotiven und Erwartungen assoziiert sind und nicht selten in Qualitätsfragen münden (Egloffstein, Kögler & Ifenthaler, 2019). Infolgedessen erscheint eine evidenz- und adressatenorientierte Entwicklung von entsprechenden Bildungsangeboten als zielführend und sinnvoll. Vor diesem Hintergrund beschäftigt sich der vorliegende Beitrag mit der übergeordneten Fragestellung, wie ein onlinebasiertes Lernangebot zu KI in der beruflichen Bildung evidenzgestützt gestaltet werden kann. Das diesem Beitrag zugrundeliegende Verbundprojekt AI_VET¹ zielte auf die Entwicklung und Implementierung von offenen Online-Kursen (i. S. v. MOOCs) aus dem Bereich der beruflichen Bildung als Reaktion auf emergente KI-bezogene Qualifikationsbedarfe in der Aus- und Weiterbildung von pädagogischen Professionals und die diesbezüglichen Leerstellen im deutschsprachigen Weiterbildungsangebot.

Der Vorgehenslogik des ADDIE (Analyze, Design, Develop, Implement, Evaluate)-Modells für Instruktionsdesign (Niegemann, 2019) folgend berichtet der Beitrag aus einer qualitativen Studie zur Exploration der Vorkenntnisse, Erwartungen und Perspektiven von Stakeholder*innen der beruflichen Bildung (Lehrpersonen, Lehramtsstudierende, betriebliches Bildungspersonal) auf ein derartiges Bildungsangebot und zeichnet erste Implementierungsansätze und korrespondierende Erkenntnisse nach. Dabei werden zunächst die Ergebnisse einer Interviewstudie mit den verschiedenen Adressatengruppen dargestellt und ihre Implikationen für die Kursentwicklung erörtert. Daran anknüpfend werden der Aufbau des Lernangebots und die zentralen Inhalte skizziert und die frühe Implementierungsphase bilanziert. Im Rahmen

¹ AI_VET: Artificial Intelligence in Vocational Education and Training; gefördert mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter den Förderkennzeichen 16DHBQP018 und 16DHBQP019

eines Ausblicks werden Potenziale und Desiderate der KI-Nutzung in der beruflichen Aus- und Weiterbildung aus Lehr- und Forschungsperspektive diskutiert.

2 Institutioneller Kontext: KI-Bildung auf dem KI-Campus

Der KI-Campus ist eine der bedeutendsten Lernplattformen für KI im deutschsprachigen Raum mit kostenlosen Online-Kursen, Videos und Podcasts zur Stärkung von KI- und Datenkompetenzen (Mah, Bernd & Dufentester, 2023). Als Teil der KI- und Digitalstrategie der deutschen Bundesregierung wird der KI-Campus durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert. Im Juli 2020 wurde die erste Beta-Version der KI-Campus Plattform gestartet. Seither haben der KI-Campus und seine institutionellen Partner*innen mehr als 50 Online-Kurse im Themenbereich KI implementiert. Gefördert wird die Entwicklung von digitalen Lernangeboten ("KI-Campus Originals") in deutscher und englischer Sprache, die durchgängig kostenlos verfügbar gemacht und offen lizenziert (CC BY-SA 4.0) werden, was sich deutlich von dem reduzierten Konzept von Offenheit auf gängigen MOOC-Plattformen (Kögler, Egloffstein & Schönberger, 2020) abhebt. Zusätzlich werden frei verfügbare digitale Angebote von Dritten zugänglich gemacht. Der KI-Campus legt einen Schwerpunkt auf technische Interoperabilität und die Kooperation mit anderen Plattformen und Initiativen. Die aktuelle Implementierung basiert dabei auf dem Learning Management System des Hasso-Plattner-Instituts, wobei ein Systemwechsel zu Moodle angestrebt wird. Die überwiegende Zahl der Lernangebote des KI-Campus sind Online-Kurse, die dem MOOC-Format entsprechen (Mah et al., 2023). Dabei handelt es sich im Wesentlichen um videobasierte Selbstlernangebote, die Quizze zur Reflektion und Leistungsfeststellung sowie zusätzliche Materialien enthalten und um kooperative Elemente ergänzt werden können (Egloffstein, 2018). Perspektivisch sollen die Angebote dabei auch selbst KI-Verfahren wie Learning Analytics und Empfehlungssysteme nutzen. Nach aktuellen Zahlen des KI-Campus haben bis Ende 2023 über 900 000 Personen die Plattform besucht. Über 40 000 Personen haben sich registriert und dabei mehr als 60 000 Kursanmeldungen vorgenommen.

Unter diesen Rahmenbedingungen wurde das Verbundprojekt AI_VET der Universitäten Mannheim (UMA) und Stuttgart (UST) initiiert, das die Entwicklung einer Serie von vier modularen offenen Online-Kursen in deutscher Sprache zum Thema KI in der beruflichen Bildung zum Ziel hatte. Adressat*innengruppen des Angebots sind Studierende in Studiengängen der beruflichen Bildung mit kaufmännischer, gewerblicher oder technischer Ausrichtung (z. B. Berufs-, Wirtschafts- und Technikpädagogik, etc.), Trainer*innen, Auszubildende, Lehrkräfte und pädagogische Professionals aus dem Bereich Aus- und Weiterbildung (inkl. zweite Phase der Lehrpersonenbildung sowie Weiterbildung/Professional Development) sowie Personen aus dem

Bildungsbereich im weiteren Sinne (über die berufliche Bildung hinaus) mit einem vertieften Interesse an KI und deren Einsatzmöglichkeiten in Bildungskontexten.

Der Entwicklungsprozess orientierte sich dabei an den Phasen des etablierten ADDIE-Rahmenmodells für Instruktionsdesign (Niegemann, 2019). Die initiale Analysephase umfasst dabei unter anderem die Analyse von Bedarfen der Adressatengruppen sowie des Kontexts des zu erstellenden Angebots (Branch, 2009). Zu diesem Zweck wurde eine Interviewstudie mit relevanten Akteur*innen der beruflichen Bildung durchgeführt.

3 Methodik und Ergebnisse der Interviewstudie zu KI in der beruflichen Bildung

3.1 Zielsetzung und Forschungsfragen

Die Studie wurde im Rahmen der Analysephase des Projekts AI_VET durchgeführt und zielte auf die Gewinnung von Aussagen zu (1) den Vorstellungen und Kenntnissen der für die Kursentwicklung relevanten Adressatengruppen zu KI, (2) dem Implementierungsstand von KI in den korrespondierenden Ausbildungsinstitutionen und -prozessen sowie (3) den Vorstellungen und Wünschen der Stakeholder*innen bezüglich der Gestaltung des Kursangebots zur Förderung von KI-Kompetenzen. Im Einzelnen wurden folgende Forschungsfragen adressiert:

- ▶ FF 1: Welche Vorstellungen und Kenntnisse bezüglich KI liegen bei Akteur*innen aus der beruflichen Bildung vor?
- ▶ FF 2: Inwieweit ist KI bereits in der beruflichen Aus- und Weiterbildungspraxis verankert?
- ▶ FF 3: Wie sollte ein Bildungsangebot zur Förderung von KI-Kompetenzen aussehen, das sich an das schulische und betriebliche Ausbildungspersonal in der beruflichen Bildung richtet?

3.2 Datenerhebung und -auswertung

Die Ergebnisse der qualitativen Studie basieren auf leitfadengestützten Interviews, die im Herbst/Winter 2020/2021 im Rahmen von universitären Lehrveranstaltungen und studentischen Abschlussarbeiten in den Masterstudiengängen für Wirtschafts- und Berufspädagogik an den Verbunduniversitäten UMA und UST durchgeführt wurden. Der Interviewleitfaden wurde durch Studierende der UMA entwickelt und erprobt und im Folgenden an der UST weiterentwickelt. Er beinhaltete ausgehend von den Forschungsfragen drei zentrale Themenbereiche und Leitfragen mit zusätzlichen Nachfragekategorien. Im Bereich A wurden die Teilnehmenden gefragt, was sie unter KI verstehen, der Bereich B enthielt die Leitfrage nach der Bedeutung von

KI für die jeweilige berufliche Tätigkeit und im Bereich C wurde erfragt, wie ein Bildungsangebot zu KI gestaltet werden müsste, damit es gewinnbringend sei.

Erste Interviews von 30 bis 60 Minuten Dauer wurden nach einer Methodenschulung der beteiligten Studierenden an der UMA durchgeführt, aufgezeichnet und anonymisiert transkribiert. Die inhaltsanalytische Auswertung erfolgte in MAXQDA im Rahmen eines deduktiv-induktiven Vorgehens (Mayring, 2022; Kuckartz & Rädiker, 2022). Dabei wurden die Daten im Rahmen einer ersten explorativen Kodierung in wechselseitiger Abstimmung in Tandems ausgewertet und im Plenum konsensuell validiert. Nach diesen ersten Schritten wurde das Vorgehen an der UST übernommen und zur Erhebung weiterer Teilstichproben weiterentwickelt. Eine Berechnung der Intercode-Übereinstimmung auf der Basis von Kodiersegmenten ergab in diesem Arbeitsschritt einen Wert von rund 84 % bei einer Segmentüberlappung von 80 %, was ausgehend von inhaltlichen Überlegungen als zufriedenstellende Reliabilität interpretiert werden kann (Rädiker & Kuckartz, 2019). Im Anschluss wurde der Kodierleitfaden noch einmal angepasst und finalisiert und diente als Basis für die Auswertungen in allen Teilstichproben, die nachfolgend dargestellt werden.

3.3 Stichprobe

Die Gesamtstichprobe der Studie betrug $N = 48$ Teilnehmende aus verschiedenen Stakeholder-Gruppen, darunter $n = 20$ Lehrpersonen, $n = 10$ angehende Lehrpersonen, $n = 12$ betriebliche Auszubildende/Akteur*innen und $n = 6$ sonstige Personen mit einschlägigem Hintergrund, zum Beispiel Bildungsverantwortliche aus Industrie- und Handels- sowie Handwerkskammern. Diese wurden in drei Teilstichproben befragt: An der UMA umfasste die erste Teilstichprobe $n = 22$ Personen, an der UST wurden in zwei weiteren Wellen jeweils 13 Personen im Abstand von wenigen Wochen befragt. Das Gros der Befragten ($n = 27$) war zwischen 20 und 29 Jahre alt, 8 Personen zwischen 30 und 39 Jahre alt und 12 Personen waren älter als 40 Jahre. Zwanzig befragte Personen gaben an, weiblichen Geschlechts zu sein, 28 Personen ordneten sich dem männlichen Geschlecht zu.

3.4 Ergebnisse

3.4.1 Kodierungen in der Übersicht

Insgesamt wurden in allen Kategorien 2 361 Kodierungen vorgenommen. Tabelle 1 zeigt einen Überblick über alle Haupt- und Unterkategorien und die entsprechenden Häufigkeiten. Auf den Bereich A zum Grundverständnis der Befragten von KI entfallen insgesamt 892 Kodierungen. Der Bereich B zur Bedeutung von KI enthält 976 Kodierungen. Der Bereich C zum Schulungsbedarf und den Vorstellungen über das Kursangebot weist 484 Kodierungen auf.

Tabelle 1: Überblick über alle Kategorien und Häufigkeiten der Kodierungen

	Kategorie	UMA	UST1	UST2	Gesamt
A	Grundverständnis von KI				
A 1	Konzeptverständnis				
A 1.1	Definitiorik	42	40	29	111
A 1.2	Systeme und Technologien	65	49	20	134
A 1.3	Anwendungsfelder	113	46	42	201
A 2	Erfahrungshintergrund				
A 2.1	Vertrautheit, Wissen, Kompetenz	132	63	44	239
A 2.2	Kompetenzerwerb	17	23	18	58
A 2.3	Qualifizierung	20	0	10	30
A 3	Allgemeine Relevanz von KI				
A 3.1	gegenwärtig	29	17	10	56
A 3.2	zukünftig	35	17	11	63
B	Bedeutung von KI				
B 1	Relevanz & Einsatz von KI				
B 1.1	gegenwärtig	138	86	38	262
B 1.2	zukünftig	106	63	59	228
B 2	Herausforderungen und Probleme von KI in der beruflichen Praxis				
B 2.1	Datenschutz	15	6	5	26
B 2.2	Akzeptanz	39	8	16	63
B 2.3	Transparenz	4	11	3	18
B 2.4	Ressourcen	41	23	28	92
B 2.5	Wissen/Kompetenzen	48	27	12	87
B 2.6	Regulierung	27	9	8	44
B 2.7	Sonstige Herausforderungen	66	47	43	156
C	Schulungs-/Unterstützungsbedarf				
C 1	Subjektive Erwartungen	60	61	49	170
C 2	Präferierte Inhalte	53	44	50	147
C 3	Learning Design	47	34	64	145
C 4	Anwendungsszenarien	11	2	9	22

	Kategorie	UMA	UST1	UST2	Gesamt
D	Restkategorie				
	Offene Kategorie	3	3	3	9
	Codesystem Summe	1111	679	571	2361

Anmerkungen: UMA = Universität Mannheim, UST = Universität Stuttgart

Innerhalb der Kategorie A zum Grundverständnis von KI verteilen sich die Kodierungen zu 50 % auf das Konzeptverständnis, zu 37 % auf den Erfahrungshintergrund und zu 13 % auf die Zuschreibung einer allgemeinen Relevanz von KI. Die Kategorie B besteht zur Hälfte aus Kodierungen zur Relevanz und dem Einsatz von KI im Arbeitsumfeld der Befragten und zur anderen Hälfte aus Kodierungen von Aussagen zu KI-assoziierten Herausforderungen und Problemen in der beruflichen Praxis. Im Bereich C, in dem es um Schulungsbedarfe und Erwartungen der Befragten an das Kursangebot ging, verteilten sich die Kodierungen zu 35 % auf subjektive Erwartungen, zu 30 % auf präferierte Kursinhalte, zu 30 % auf das Learning Design und zu 5 % auf konkrete Anwendungsszenarien. Nachfolgend werden die Ergebnisse in den drei Hauptkategorien inhaltlich verdichtet dargestellt und im Anschluss zusammengefasst.

3.4.2 Grundverständnis von KI

Bei der Einstiegsfrage nach dem Grundverständnis der Befragten in Bezug auf KI zeigt sich ein sehr heterogenes Bild, das insgesamt ein wenig einheitliches und eher oberflächliches Verständnis von KI dokumentiert. Die Spanne der Definitionen reicht von „Künstliche Intelligenz ist Digitalisierung“ [EIF02B, Pos. 12] über „Künstliche Intelligenz ist irgendetwas, wo man auf einen Knopf drückt“ [EIF01A, Pos. 15] bis hin zu vergleichsweise präziseren Definitionen wie „Unter Künstlicher Intelligenz verstehe ich einen programmierten Algorithmus in einem Computer, der eben mit Daten, die ihm gefüttert werden, bestimmte Entscheidungsstränge basteln kann, sodass er dann mit neuen Informationen dann eigene, also eigene - in Anführungszeichen - Entscheidungen treffen kann“ [EIF03A, Pos. 8].

Oftmals enthalten die Aussagen der Teilnehmenden einen Bezug zu menschlichem Denken und Handeln, bisweilen verschwimmen die Grenzen zwischen KI und Machine Learning - „Gut, also Künstliche Intelligenz ist irgendwo Maschinen gesteuerte oder Maschinen unterstützter Prozess, der letztendlich versucht sag ich mal so diese Gehirn-technischen Leistung des Menschen zu kopieren und im Idealfall sogar weiterzuentwickeln“ [EIF05B, Pos. 11]. Besonders häufig werden unterschiedliche Anwendungsfelder von KI aus den Bereichen Alltagsleben (z. B. Sprachassistent, Recommender Systems), Bildung (z. B. Lernplattformen), Industrie und Wirtschaft

(v. a. Automotive) und Gesellschaft (z. B. Gesichtserkennung) angesprochen. Die Erfahrungshintergründe der Befragten entstammen im Wesentlichen dem privaten Bereich. Der Kompetenzerwerb erfolgt dabei vor allem auf informellem Wege (z. B. über Learning by doing, Medien, unter Kolleg*innen, auf Messen) und weniger über formale Qualifikationsangebote (z. B. universitäre Lehre für angehende Lehrpersonen). Ungeachtet der Heterogenität der Perspektiven auf KI herrscht in der Stichprobe ein breiter Konsens hinsichtlich der zukünftig wachsenden Bedeutung von KI für das Individuum und die Gesellschaft.

3.4.3 Bedeutung und Verankerung von KI im Praxisfeld

In der Stichprobe gibt es insbesondere bei (angehenden) Lehrpersonen kaum Berührungspunkte mit KI in der aktuellen beruflichen Praxis - „Sagen wir mal so ... Es wird vielleicht ein bisschen angeschnitten. Das Thema Big Data ein Stück weit. Aber so die klassischen KI-Techniken, dass man da über neuronale Netze oder so etwas im Detail spricht – wenn, dann extremst am Rande“ [EIF08A, Pos. 15] - wohingegen die Thematik im betrieblichen Umfeld eine größere Rolle spielt und in ersten Projekten implementiert wird („Also bei uns ist es im Prinzip so, dass wir im Moment relativ viel stark schauen, was bei KI an Möglichkeiten besteht“ [EIF03B, Pos. 16]).

Bei allen Befragten herrscht ein breiter Konsens bezüglich der zunehmenden Relevanz von KI. Es wird zukünftig ein deutlich häufigerer Einsatz von KI im beruflichen Kontext erwartet. Insbesondere das betriebliche Ausbildungspersonal benennt in diesem Zusammenhang viele konkrete Szenarien, während Lehrkräfte sich dagegen realistisch bis eher skeptisch äußern, was die Umsetzung angeht („Inwieweit in den Schulen tatsächlich KI angewendet wird in Zukunft, da bin ich eher kritisch, muss ich zugeben. Weil wenn man mal betrachtet, wie die Schulen oder Einrichtungen digitalisiert sind, dann sieht es ja auch nicht so gut aus. Also könnte ich mir nicht vorstellen, dass es da in zehn Jahren so einen riesigen Sprung gibt. Also zumindest nicht flächendeckend, also vielleicht an einzelnen Stellen, aber ja, genau“ [E01M_0912]).

Besonders differenziert werden unterschiedliche Herausforderungen im Zusammenhang mit KI angesprochen (486 Kodierungen, s. Abb. 1) – dabei spielen insbesondere Sorgen bezüglich fehlender Ressourcen und mangelnden Wissens bzw. mangelnder Kompetenzen sowie Akzeptanzfragen eine übergeordnete Rolle. Das weite Spektrum der Antworten führt zu einer stark ausgeprägten Kategorie der ‚Sonstigen Herausforderungen‘. Hier werden bspw. Befürchtungen hinsichtlich Autonomieverlusten, dem Entstehen eines Überwachungsstaates oder dem Ersetzt-Werden durch Maschinen sowie ethische Bedenken geäußert.

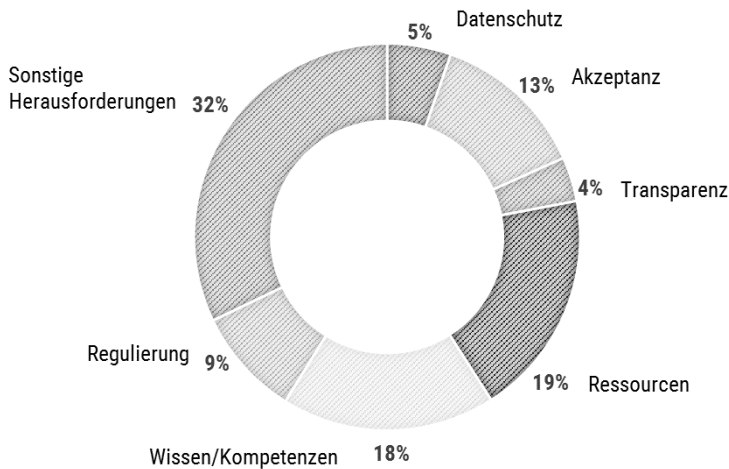


Abbildung 1: Herausforderungen im Zusammenhang mit KI

3.4.4 Erwartungen an das Bildungsangebot und Schulungsbedarf

Die durch die Befragten geäußerten Erwartungen an das Bildungsangebot beziehen sich inhaltlich auf (1) die Vermittlung von Basics und einem Grundverständnis zu KI, (2) konkrete Anwendungsbeispiele und Use-Cases und (3) die Auseinandersetzung mit Chancen, Risiken und Limitationen von KI in beruflichen Anwendungskontexten. Als angemessener Zeitaufwand für eine KI-bezogene Schulung werden über alle Stakeholder-Gruppen hinweg 1 bis 2 Stunden pro Woche angesehen. Hinsichtlich des Learning Designs wünschen sich die Befragten in erster Linie einen niedrigschwelligen Online-Kurs mit qualitativ hochwertigen und ansprechend umgesetzten Videos. Weiterhin werden Interaktion und Austausch mit anderen Teilnehmenden, etwa in Chats und Foren, als sinnvoll und wichtig erachtet. Auf die Bedeutung persönlicher Ansprechpartner*innen sowie einer aktiven Betreuung wird hingewiesen.

3.5 Zusammenfassung und Implikationen

In der Gesamtschau ergibt sich aus der Interviewstudie ein differenziertes Befundbild, das für die evidenzgestützte Kursentwicklung genutzt werden kann. In Bezug auf Forschungsfrage eins (Vorstellungen und Kenntnisse) zeichnet sich einerseits ein breiter Konsens der Befragten bezüglich der Relevanz von KI im Kontext der beruflichen Bildung ab (vgl. Roppertz, 2021; Ehlers, Lindner & Rauch, 2023). Andererseits offenbaren sich eher diffuse Vorstellungen bezüglich des Begriffsumfangs und kor-

respondierender Einsatzfelder von KI. Dieser Umstand verweist auf den grundlegenden Bedarf für ein niedrigschwelliges, breit angelegtes Bildungsangebot, das ein Grundverständnis für KI im Allgemeinen fördert und auch bei unterschiedlichen Vorwissens- und Kompetenzständen (vgl. Rott, Lao, Petridou & Schmidt-Hertha, 2022) eingesetzt werden kann. Detailliertere Implikationen zur inhaltlichen Gestaltung eines Lernangebots (im Sinne einer Wissens- und Aufgabenanalyse gemäß ADDIE) können angesichts der diffusen Ausgangslage aber nicht oder nur sehr bedingt abgeleitet werden.

Hinsichtlich der zweiten Forschungsfrage (Verankerung in der Aus- und Weiterbildungspraxis) zeigen sich deutliche Unterschiede zwischen den Adressat*innengruppen. Lediglich im betrieblichen Kontext gibt es erste konkrete Anwendungsbezüge von KI, in der universitären Ausbildung und im schulischen Kontext spielte die Thematik zum Zeitpunkt der Befragung eine untergeordnete Rolle. Mit Blick auf die Unterschiede zwischen schulischen und betrieblichen Ausbildungsverantwortlichen erscheint eine differenzierte Zielgruppenansprache wünschenswert. Für die Kursentwicklung wäre hierzu eine modulare, adressatengerechte Gestaltung mit möglichst vielfältigen Anwendungsbeispielen anzustreben, um unterschiedliche Weiterbildungsbedarfe adäquat adressieren zu können.

Bezüglich Forschungsfrage drei (Anforderungen an ein KI-Bildungsangebot) ist festzustellen, dass die Herausforderungen auf dem Weg zur Implementierung von KI als Inhalt (Lernen über KI) und Methode (Lernen mit KI) in der beruflichen Aus- und Weiterbildung immens sind und von den Befragten auch differenziert wahrgenommen werden. Hier ergibt sich der Bedarf für eine ausführliche Beschäftigung mit den rechtlichen, regulatorischen und institutionellen Herausforderungen im zu erstellenden Lernangebot. Der Wunsch der Befragten nach Interaktion mit anderen Teilnehmenden und Individualbetreuung durch Kursverantwortliche knüpft an existierende Befunde hinsichtlich der didaktischen Qualität in MOOCs an (Egloffstein et al., 2019) und stellt gleichzeitig eine Herausforderung im Hinblick auf die Skalierbarkeit des Angebotes dar (Kasch, van Rosmalen & Kalz, 2021).

Ungeachtet dieser Erkenntnisse ist einschränkend festzuhalten, dass in der Studie eine reine Gelegenheitsstichprobe untersucht worden ist. Die erste Datenauswertung erfolgte deskriptiv in einem niedrig-inferenten Verfahren. Weitergehende Analysen könnten bspw. eine zusätzliche Auswertungsebene durch Verdichtung des Materials erschließen und eine qualitative Bewertung der KI-Kennnisse ermöglichen. Die Stichprobe würde zumindest Tendenzaussagen zu Zusammenhängen innerhalb der Themenbereiche oder zu Unterschieden zwischen den Stakeholder-Gruppen ermöglichen.

4 Design und Implementierung der Kursreihe AI_VET

Auf Basis der Ergebnisse der Interviewstudie und weiterführenden theoretischen Überlegungen wurde im AI_VET-Projektverbund zusammen mit dem KI-Campus das Konzept für eine Kursreihe zum Themenkomplex KI in der beruflichen Bildung erarbeitet. Für einen modularen, zielgruppenadäquaten Zugang wurden vier Online-Kurse mit unterschiedlichen thematischen Zuschnitten (ein Einführungskurs, drei Vertiefungskurse) implementiert.

4.1 Kompetenzziele

Die Ergebnisse der Interviewstudie machen deutlich, dass in der Zielgruppe grundsätzliche Weiterbildungsbedarfe hinsichtlich KI bestehen. Auf dieser Basis und im Einklang mit den Vorgaben des KI-Campus wurde daher der Erwerb von KI-Grundlagenkompetenzen unter Berücksichtigung der instrumentellen (KI als Lernwerkzeug) wie auch der inhaltlichen (KI als Lerngegenstand) Perspektive als Zielstellung der Kursreihe AI_VET formuliert. Die übergreifenden Kompetenzziele dazu lauten:

Die Lernenden...

- ▶ kennen grundlegende Konzepte der KI und deren Anwendungsmöglichkeiten im Kontext der Berufsbildung;
- ▶ verstehen grundlegende Funktionsprinzipien von KI-Anwendungen im Kontext der Berufsbildung;
- ▶ können Einsatzbereiche, Möglichkeiten und Grenzen von KI-Technologien im Bildungsbereich einschätzen;
- ▶ können neue Entwicklungen aus dem KI-Bereich auf deren Potenziale für den Bildungsbereich analysieren;
- ▶ können Einsatzszenarien für KI-Technologie im Bildungsbereich abstecken und KI-getriebene pädagogisch-didaktische Interventionen entwerfen;
- ▶ können Anforderungen an KI-Systeme für den beruflichen Bildungsbereich formulieren und Richtlinien zu deren Anwendung entwerfen.

Diese übergreifenden Kompetenzziele wurden anschließend in Learning Outcomes für alle vier Kurse ausdifferenziert. Festzuhalten ist dabei, dass die Kompetenzziele in Anknüpfung an die Interviewstudie eher allgemein formuliert wurden, um dem heterogenen Wissensstand der Zielgruppe gerecht zu werden.

4.2 Didaktisches Design

Das didaktische Design der AI_VET-Kurse orientiert sich stark an gängigen MOOC-Modellen, wobei der didaktische Handlungsspielraum durch organisatorische und finanzielle Rahmenbedingungen sowie die Möglichkeiten der verwendeten Plattform vorgegeben wird (Blackmon & Major, 2017). So wurden die AI_VET-Kurse als

Selbstlernumgebungen geplant, die ohne zeitliche und örtliche Restriktionen durchlaufen werden können. Durch den Verzicht auf eine zeitliche Taktung (keine fixen Start- und Enddaten, keine feste Laufzeit) werden allerdings die Möglichkeiten für Interaktionen zwischen den Lernenden sowie für tutorielle Betreuung eingeschränkt, was den in der Interviewstudie geäußerten Vorstellungen zuwiderläuft. Der modulare Aufbau ermöglicht dafür eine hohe didaktische und organisatorische Flexibilität (Egloffstein & Ifenthaler, 2023) mit unterschiedlichen Skalierungsmöglichkeiten durch die ‚stand-alone‘ Nutzung von einzelnen Kursen und Kursbausteinen als Open Educational Resources (Stracke, Downes, Conole & Burgos, 2019). Dadurch wird bspw. die Einbindung der Kursreihe in unterschiedliche Blended Learning-Szenarios ermöglicht, die sich an unterschiedliche Vorwissens- und Kompetenzstände richten können. Darüber hinaus wird das traditionelle MOOC-Modell um eine zweite didaktischen Ebene in Anlehnung an den Goal Based Scenario-Ansatz (Schank, Fano, Bell & Jona, 1994) erweitert. Dabei werden eine durchgängige Rahmengeschichte (die sog. Cover Story) sowie explizite Handlungsziele (sog. Missionen) in die Aufgabenstellungen der einzelnen Kursmodule integriert, um eine Kontextualisierung der Inhalte zu erreichen und die Problemorientierung zu fördern.

Das videobasierte Kursformat wird durch Quizze in unterschiedlichen Formaten und Reflexionsübungen sowie durch textuelle Inhalte ergänzt. Neben instruktionalen Lehrvideos werden dabei zusätzliche Videos mit Inhalts- und Praxisexpert*innen (z. B. von der Mercedes-Benz AG) angeboten, um den Kontext noch weiter zu vertiefen und die in der Studie angesprochenen Praxisbezüge herzustellen. Textmaterialien (Inhaltsseiten, Videotranskripte) stehen zusätzlich als Download zur Verfügung. Der ungefähre Zeitaufwand für die Bearbeitung der Kursreihe beträgt 150 Stunden, wovon jeweils 25 Stunden auf die Kurse AI_VET 1 und 4 sowie jeweils 50 Stunden auf die Kurse AI_VET 2 und 3 entfallen.

Die Leistungsbeurteilung findet in gesonderten Quizzen nach jedem Inhaltsabschnitt statt. Gemäß den Zertifikatsregeln des KI-Campus erhalten Lernende einen Leistungsnachweis, wenn sie in diesen Prüfungen mind. 60 % der möglichen Punkte erreichen. Eine Teilnahmebestätigung erhält, wer auf mindestens 50 % der Kursmaterialien zugegriffen hat. Tabelle 2 zeigt die Inhalte der vier AI_VET-Kurse im Überblick.

Tabelle 2: Inhalte der AI_VET-Kurse

AI_VET 1 KI im Kontext von Lehren & Lernen	AI_VET 2 Learning Analytics	AI_VET 3 KI als Inhalt der beruflichen Bildung	AI_VET 4 KI als Werkzeug in der beruflichen Bildung
(Digitales) Lehren und Lernen	Begriffsverständnis Learning Analytics (LA)	Transformation durch Digitalisierung und KI	Status Quo der Digitalisierung in der Berufsbildung
KI im Bildungskontext	Forschungsstand zu LA	KI und Arbeitswelt KI in der Produktion	Voraussetzungen der Nutzung von KI
Potenziale von KI im Bildungs- kontext	Implementation von LA	KI in weiteren Unternehmensbereichen	KI mit Hardware- und Realweltbezug
Herausforderun- gen von KI im Bildungskontext	Herausforderungen bei Datenanalyse	Ethische Aspekte von KI Regulatorische Aspekte von KI	KI in Softwaresystemen und Programmen
	Herausforderung Datenschutz	Berufsbildung 4.0 Veränderte	
	Change Management	Qualifikationsanforderungen	
	Handlungsempfehlungen		

4.3 Implementierung und erste Erfahrungen

Im Projektablauf wurden nach der Analysephase auch die weiteren Phasen des ADDIE-Prozesses durch den Einbezug von Stakeholdern (hier: Lehramtsstudierende) unterstützt. Dies ermöglichte kollaborative Ideenfindungsprozesse, Prototyping und eine partizipative Gestaltung der Kursangebote. Im Rahmen der Studiengänge Wirtschaftspädagogik (an der UMA) sowie Berufspädagogik (an der UST) wurden drei Projektseminare (PS) durchgeführt, die die didaktische Gestaltung von Online-Angeboten zum Gegenstand hatten. Im PS1 (UMA, Sommersemester 2021, 17 Teilnehmende) analysierten die Studierenden die AI_VET-Storyboards auf Konsistenz, Schwierigkeitsniveau, didaktisch angemessenen Medieneinsatz usw. Weiterhin wurden Entwürfe für die wesentlichen Elemente der Goal Based Scenarios entwickelt, also etwa für die Coverstory, die Mission und mögliche Assessments. Im PS2 (UST, Sommersemester 2021, 5 Teilnehmende) analysierten die Studierenden Videoskripte und Storyboards und entwickelten konkrete Ideen für die Kursevaluation. Beide Projektseminare zur Unterstützung der Design- und Entwicklungsphasen (-DD-) wurden pandemiebedingt vollständig online durchgeführt. Projektseminar 3 (UST, Wintersemester 2021/22, 10 Teilnehmende) war ein hybrider Kurs mit dem Schwerpunkt auf den Implementierungs- und Evaluierungsphasen (IE). Die Studierenden erarbeiteten Konzepte für die summative Evaluation von AI_VET und testeten ihre Konzepte anhand von bestehenden Kursmaterialien.

Ab Januar 2022 wurden die vier Kurse sukzessive live geschaltet. Bis Januar 2024 stellt sich die Kursnutzung folgendermaßen dar (Tabelle 3):

Tabelle 3: Nutzung der AI_VET-Kurse

Nutzungsindikator	AI_VET 1	AI_VET 2	AI_VET 3	AI_VET 4
Kurseinschreibungen	537	204	179	194
No-Shows	85	42	47	41
Teilnahmebescheinigungen	155	57	36	45
Leistungsnachweise	82	40	21	20

Anmerkungen: No-Shows = keine weitere Aktivität nach initialer Anmeldung

Die vergleichsweise geringe Anzahl an Leistungsnachweisen im Verhältnis zur Zahl der aktiv Lernenden deutet darauf hin, dass die Lernenden die AI_VET-Angebote in der Mehrzahl nicht als eigenständige, abzuschließende Kurse, sondern eher als modulare Inhaltsbausteine bspw. innerhalb von Blended Learning-Szenarios wahrgenommen haben. Die vermeintlichen „Abschlussquoten“ bewegen sich dabei im Einklang mit den einschlägigen Ergebnissen aus der MOOC-Forschung um einen Wert von etwa 20 % (vgl. Jordan, 2015).

Zum Ende des Kurses wurden die Teilnehmenden per Online-Fragebogen um Einschätzungen zu ihren Kenntnissen und ihrem Lernprozess gebeten. Tabelle 4 zeigt erste Ergebnisse.

Tabelle 4: Selbsteinschätzung des Lernerfolgs in den AI_VET-Kursen

Item	AI_VET 1 n = 68	AI_VET 2 n = 20	AI_VET 3 n = 9	AI_VET 4 n = 10
Kenntnisse vor dem Kurs				
sehr umfangreich	5 %	0 %	11 %	0 %
eher umfangreich	15 %	5 %	11 %	10 %
mittelmäßig	33 %	35 %	33 %	50 %
eher gering	36 %	45 %	44 %	40 %
sehr gering	11 %	15 %	0 %	0 %
Kenntnisse nach dem Kurs				
verbessert	67 %	75 %	78 %	90 %
gleich	33 %	25 %	22 %	10 %
Erreichen der eigenen Lernziele				
vollständig	15 %	15 %	44 %	40 %
größtenteils	51 %	75 %	33 %	50 %
teilweise	31 %	10 %	11 %	0 %
eher nicht	1 %	0 %	11 %	0 %
überhaupt nicht	1 %	0 %	0 %	0 %
Weiterempfehlung				
ja	93 %	95 %	78 %	90 %
nein	7 %	5 %	22 %	10 %

Anmerkungen: Gerundete relative Häufigkeiten, fehlende Werte nicht dargestellt.

5 Diskussion und Ausblick

Der vorliegende Beitrag war dem Ziel gewidmet, den evidenzorientierten Entwicklungs- und Implementierungspfad von offenen, onlinebasierten Bildungsangeboten zu KI in der beruflichen Bildung nachzuzeichnen. Im Fokus stand dabei besonders die Analysephase im Sinne des ADDIE-Modells des Instruktionsdesigns. Aus N = 48 Interviews mit Vertreter*innen wesentlicher Adressatengruppen konnten Erkenntnisse zum KI-bezogenen Vorwissen und den Erwartungen an ein Kursangebot zur Förderung von KI-Kompetenzen sowie Hinweise zu aktuellen Anwendungsszenarien in berufsbildenden und beruflichen Kontexten gewonnen werden. Die Erkenntnisse aus den Interviews dienten dabei nicht primär zur direkten Ableitung von Kursinhalten (diese erfolgte mehrheitlich theoriegeleitet und literaturbasiert), sondern waren

wertvolle Impulse für die didaktisch-methodische Kursgestaltung etwa mit Blick auf die Formulierung von Kompetenzzielen, die Einbindung von Implementierungsbeispielen in den gewerblich-technischen und kaufmännischen Zieldomänen sowie den Anspruchsgehalt der lernzielorientierten Assessments. Die generierten Ergebnisse weisen insgesamt eine gute Anschlussfähigkeit an bestehende Befunde auf. Dabei fallen die gemessen an der Bedeutung des Gegenstandsbereichs relativ überschaubaren KI-Kenntnisse insbesondere von Lehrkräften sowie das zum Erhebungszeitpunkt vergleichsweise spärliche Anwendungswissen in der Stichprobe auf. Dieser Befund ist allerdings mit Blick auf die dynamischen Entwicklungen rund um generative KI-Tools spätestens seit Ende 2022 insofern zu relativieren, als gerade die Niedrigschwelligkeit (im Sinne einer Nutzbarkeit ohne nennenswerte KI-Kenntnisse) und allgemeine Verfügbarkeit von KI zu einer intensiveren Anwendung nicht zuletzt in beruflichen Bildungssettings führen kann (vgl. Jacobs & Jaschke, 2023). Ungeachtet dessen stellte die entwickelte Kursreihe zum Zeitpunkt der Implementierung eines der ersten deutschsprachigen Angebote in der beruflichen Bildung dar und löste somit den Anspruch ein, ein Grundverständnis zu KI als Inhalt und Methode der beruflichen Bildung zu etablieren. Angesichts der aktuellen Entwicklungsdynamik sind allerdings nicht nur laufende Aktualisierungen der Kursinhalte notwendig. Vielmehr ist im Sinne eines Entwicklungspfades der KI-Bildung eine weitere Ausdifferenzierung von Bildungsangeboten hinsichtlich kontextspezifisch benötigter Kompetenzen erforderlich und wird etwa in Gestalt von ersten Kursangeboten zur Bildungsdatenkompetenz für Lehrkräfte (vgl. Projekt BDKomp der UMA) bereits implementiert.

Insgesamt stellen sich nach der Implementierung der Kursreihe und ersten Erkenntnissen zur Kursnutzung und Lernzielerreichung unterschiedliche Fragen, die sowohl die Lehr- als auch die Forschungsperspektive betreffen. Auf dem Weg zu einer nachhaltigen Integration der Lernangebote im Hochschulkontext sind aus Lehrperspektive curriculare sowie medien- und hochschuldidaktische Aspekte zu klären, die im Rahmen einer initialen Kooperation mit der Universität Leipzig adressiert werden. Hier werden in mittelfristiger Perspektive verschiedene Szenarien der Einbindung der Kursreihe in die akademische Ausbildung erprobt. Langfristig ist eine Diskussion bezüglich der Anpassung der disziplinären Rahmenvorgaben (BWP, 2014) wünschenswert, um den Stellenwert von KI als Inhalt und Methode breit in den berufs- und wirtschaftspädagogischen Hochschulcurricula zu verankern.

Aus der Forschungsperspektive heraus lässt sich das Projekt AI_VET in eine Reihe von Ansätzen einordnen, die die Förderung eines grundlegenden Verständnisses von KI, einer sog. KI-Literacy zum Ziel haben (Laupichler, Aster, Schirch & Raupach, 2022). Da es nach wie vor an weithin akzeptierten KI-Kompetenzmodellen mangelt

und die technische Entwicklung zügig voranschreitet, ist die Frage, welche KI-Kompetenzen Nicht-Experten erwerben sollten, alles andere als trivial. Vor diesem Hintergrund zeichnen sich für Lernangebote zu KI-Literacy im Wesentlichen zwei Entwicklungspfade ab: zum einen in Richtung einer Spezialisierung, die auf den Erwerb von domänenspezifischen, meist technologieorientierten KI-Kompetenzen (Ng. et al., 2023) abzielt, die weit über grundlegende Fähigkeiten und ein Bewusstsein für die Technologie und ihre Wirkungen hinausgehen. Zum anderen in Richtung einer transdisziplinären KI-Bildung, die KI als Grundlage in allen Disziplinen und Domänen sieht (Aliabadi, Singh & Wilson, 2023). Auf lange Sicht muss vor allem die zweite Perspektive angesprochen werden, wenn sich die Potenziale von Künstlicher Intelligenz gesamtgesellschaftlich entfalten sollen. Bildungsangebote wie AI_VET stellen einen ersten, nicht unwesentlichen Schritt auf diesem Wege dar.

Literatur

- Aliabadi, R., Singh, A. & Wilson, E. (2023). Transdisciplinary AI education: The confluence of curricular and community needs in the instruction of artificial intelligence. In T. Schlippe, E. C. K. Cheng & T. Wang (Eds.), *Artificial Intelligence in Education Technologies: New Development and Innovative Practices*. Proceedings AIET 2023 (pp. 137-151). Singapore: Springer Nature.
- Attwell, G., Bekiaridis, G., Deitmer, L., Perini, M., Roppertz, S. & Tutlys, V. (2020). Artificial Intelligence in policies, processes and practices of vocational education and training. ITB-Forschungsbericht 71. <https://www.doi.org/10.26092/elib/307>
- Blackmon, S. J. & Major, C. H. (2017). Wherefore art thou MOOC? Defining massive open online courses. *Online Learning Journal*, 21 (4), 195-221. <https://www.doi.org/10.24059/olj.v21i4.1272>
- Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. New York, NY: Springer. <https://www.doi.org/10.1007/978-0-387-09506-6>
- BWP (2014). Basiscurriculum für das universitäre Studienfach Berufs- und Wirtschaftspädagogik im Rahmen berufs- und wirtschaftspädagogischer Studiengänge. Verfügbar unter: www.dgfe.de/fileadmin/OrdnerRedakteure/Sektionen/Sek07_BerWiP/2014_Basiscurriculum_BWP.pdf [31.01.2024].
- Chiu, T. K. F. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1-17. <https://www.doi.org/10.1080/10494820.2023.2253861>
- Ebner, M., Schön, S. & Braun, C. (2019). Mehr als nur ein MOOC: Sieben Lehr- und Lernszenarien zur Nutzung von MOOCs in der Hochschulelehre und anderen Bildungsbereichen. In J. Hafer, M. Mauch & M. Schumann (Hrsg.), *Teilhabe in der digitalen Bildungswelt*. GMW Proceedings 2019 (S. 138-149). Münster: Waxmann.
- Egloffstein, M. (2018). Massive open online courses in digital workplace learning: Current state and future perspectives. In D. Ifenthaler (Ed.), *Digital workplace learning: Bridging formal and informal learning with digital technologies* (pp. 149-166). Cham: Springer. https://www.doi.org/10.1007/978-3-319-46215-8_9
- Egloffstein, M. & Ifenthaler, D. (2023). Modularizing MOOCs for professional learning: Prospects and challenges for instructional design. In *Proceedings 2023 IEEE Learning with MOOCs* (pp. 1-4). <https://www.doi.org/10.1109/LWMOOC58322.2023.10305982>
- Egloffstein, M., Kögler, K. & Ifenthaler, D. (2019). Instructional quality of business MOOCs: Indicators and initial findings. *Online Learning Journal*, 23 (4), 85-105. <https://www.doi.org/10.24059/olj.v23i4.2091>
- Ehlers, U.-D., Lindner, M. & Rauch, E. (2023). *AIComp – Future skills for a world shaped by AI*. Karlsruhe. Verfügbar unter: www.ai-comp.org [31.01.2024].

- Ertl, H. & Seifried, J. (2021). Aus-wirkungen künstlicher Intelligenz auf Arbeits-prozesse und Qualifikations-anforderungen in verschiedenen Berufsfeldern – Ein Kommentar zu Teil A des Beihefts. Zeitschrift für Berufs- und Wirtschaftspädagogik, Beiheft, 31, 153-160.
- Europäische Kommission (2013). Opening up education: Innovative teaching and learning for all through new technologies and open educational resources. Brüssel: Europäische Kommission. Verfügbar unter: eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52013DC0654 [31.01.2024]
- Ifenthaler, D. (2020). Learning Analytics im Hochschulkontext – Potenziale aus Sicht von Stakeholdern, Datenschutz und Handlungsempfehlungen. In R. A. Fürst (Hrsg.), Digitale Bildung und Künstliche Intelligenz in Deutschland. Nachhaltige Wettbewerbsfähigkeit und Zukunftsagenda (S. 519-535). Wiesbaden: Springer
- Jacobs, S. & Jaschke, S. (2023): Large Language Models in der Berufsausbildung von IT-Fachkräften. INFOS 2023 - Informatikunterricht zwischen Aktualität und Zeitlosigkeit. Bonn: Gesellschaft für Informatik. <https://www.doi.org/10.18420/infos2023-017>
- Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. The International Review of Research in Open and Distributed Learning, 16 (3). <https://www.doi.org/10.19173/irrodl.v16i3.2112>
- Kasch, J., Van Rosmalen, P. & Kalz, M. (2021). Educational scalability in MOOCs: Analysing instructional designs to find best practices. Computers & Education, 161, 104054, 1–12. <https://www.doi.org/10.1016/j.compedu.2020.104054>
- Kögler, K., Egloffstein, M. & Schönberger, B. (2020). Openness in MOOCs for training and professional development – An exploration of entry and participation barriers. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), Vocational education and training in the age of digitization: Challenges and opportunities (pp. 205–223). Opladen: Barbara Budrich. <https://www.doi.org/10.3224/84742432>
- Kuckartz, U. & Rädiker, S. (2022). Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. (5., überarb. Aufl.). Weinheim: Beltz Juventa.
- Laupichler, M. C., Aster, A., Schirch, J. & Raupach, T. (2022). Artificial intelligence literacy in higher and adult education: A scoping literature review. Computers and Education: Artificial Intelligence, 3, 100101. <https://www.doi.org/10.1016/j.caeai.2022.100101>
- Littenberg-Tobias, J. & Reich, J. (2020). Evaluating access, quality, and equity in online learning: A case study of a MOOC-based blended professional degree. The Internet and Higher Education, 47, 10759. <https://www.doi.org/10.1016/j.iheduc.2020.100759>
- Mah, D. K., Bernd, M. & Dufentester, C. (2023). Evaluation und Maintenance von Online-Kursen zum Thema Künstliche Intelligenz Artikelinhalt. Zeitschrift für Hochschulentwicklung, 18 (1), 37-56. Verfügbar unter: www.zfhe.at/index.php/zfhe/article/view/1735 [31.01.2024].
- Mah, D. K., Rampelt, F., Dufentester, C., Bernd, M., Gamst, C. & Weygandt, B. (2020). Digitale Lernangebote zum Thema Künstliche Intelligenz. Überblicksstudie zu kostenlosen Online-Kursen auf deutschen Lernplattformen. Verfügbar unter: ki-campus.org/sites/default/files/2020-11/Studie_Digitale_Lernangebote_zum_Thema_KI.pdf [31.01.2024].
- Mayring, P. (2022). Qualitative Inhaltsanalyse: Grundlagen und Techniken. (13., aktual. Aufl.). Weinheim: Beltz.
- Ng, D. T. K., Lee, M., Tan, R. J. Y., Hu, X., Downie, J. S. & Chu, S. K. W. (2023). A review of AI teaching and learning from 2000 to 2020. Education and Information Technologies, 28, 8445-8501. <https://www.doi.org/10.1007/s10639-022-11491-w>
- Niegemann, H. (2019). Instructional Design. In: Niegemann, H., Weinberger, A. (Hrsg.), Handbuch Bildungstechnologie. Heidelberg: Springer. https://www.doi.org/10.1007/978-3-662-54373-3_7-1
- Rädiker, S. & Kuckartz, U. (2019). Analyse qualitativer Daten mit MAXQDA. Wiesbaden: Springer VS. https://www.doi.org/10.1007/978-3-658-22095-2_19
- Roppertz, S. (2021). Die Rolle und Bedeutung von Künstlicher Intelligenz in der Berufsausbildung – Implikationen für angehende Berufs- und Wirtschaftspädagog*innen. bwp@ Berufs- und Wirtschaftspädagogik - online 40. Verfügbar unter: www.bwpat.de/ausgabe40/roppertz_bwpat40.pdf [31.01.2024].

- Rott, K. J., Lao, L., Petridou, E. & Schmidt-Hertha, B. (2022). Needs and requirements for an additional AI qualification during dual vocational training: Results from studies of apprentices and teachers. *Computers and Education: Artificial Intelligence*. <https://www.doi.org/10.1016/j.caeai.2022.100102>
- Schank, R. C., Fano, A., Bell, B. & Jona, M. (1994). The design of goal-based scenarios. *Journal of the Learning Sciences*, 3 (4), 305-345.
- Serth, S., Staubitz, T., van Elten, M. & Meinel C. (2022). Measuring the effects of course modularizations in online courses for life-long learners. *Frontiers in Education*, 7, 1008545. <https://www.doi.org/10.3389/educ.2022.100854>
- Seufert, S. (2020). Digitale Kompetenzen in der kaufmännischen Domäne im Kontext der Berufsbildung in der Schweiz. In B. Ziegler & R. Tenberg (Hrsg.), *Berichte zur beruflichen Bildung. Berufsbildung 4.0: Steht die berufliche Bildung vor einem Umbruch?* (S. 56-73). Opladen: Verlag Barbara Budrich.
- Seufert, S., Guggemos, J. & Ifenthaler, D. (2021). Zukunft der Arbeit mit intelligenten Maschinen: Implikationen der Künstlichen Intelligenz für die Berufsbildung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, Beiheft 31, 9-27.
- Söllner, M., Janson, A., Rietsche, R. & Thiel de Gafenco, M. (2021). Individualisierung in der beruflichen Bildung durch Hybrid Intelligence. *Potentiale und Grenzen. Zeitschrift für Berufs- und Wirtschaftspädagogik*, Beiheft 31, 163-181.
- Spöttl, G. & Windelband, L. (2019). Industrie 4.0 - Neugestaltung industrieller Prozesse und Konsequenzen für die Berufsbildung. In G. Spöttl & L. Windelband (Hrsg.), *Industrie 4.0 - Risiken und Chancen für die Berufsbildung*. (S. 239-254, 2. überarb. Aufl.). Bielefeld: wbv.
- Stracke, C. M., Downes, S., Conole, G. & Burgos, D. (2019). Are MOOCs Open Educational Resources? A literature review on history, definitions, and typologies of OER and MOOCs. *Open Praxis*, 11 (4), 331-341. <https://www.doi.org/10.5944/openpraxis.11.4.1010>
- Wilbers, K. (2020). Data-Science und KI in der wirtschaftswissenschaftlichen Ausbildung. *VLB-Akzente*, 29 (11), 11.
- Wittpahl, V. (Hrsg.). (2019). *Künstliche Intelligenz. Technologie | Anwendung | Gesellschaft*. Berlin: Springer Vieweg.

Autor*innen

Dr. Marc Egloffstein, Universität Mannheim, Area Wirtschaftspädagogik, Lehrstuhl Technologiebasiertes Instruktionsdesign

Prof. Dr. Kristina Kögler, Universität Stuttgart, Abteilung Berufs-, Wirtschafts- und Technikpädagogik, Professur für Berufspädagogik

Prof. Dr. Dirk Ifenthaler, Universität Mannheim, Area Wirtschaftspädagogik, Lehrstuhl Technologiebasiertes Instruktionsdesign, Curtin University, Australien
 UNESCO Deputy Chair on Data Science in Higher Education Learning and Teaching
 Korrespondenz an: egloffstein@uni-mannheim.de

Julia Pargmann, Anna Leube, Florian Berding, Elisabeth Riebenbauer, Karin Rebmann, Andreas Slopinski & Michael Gillert

Die elektronisch-didaktische Assistenz (EDDA) zur Unterrichtsplanung auf Basis künstlicher Intelligenz: Funktionsweise, Anwendungsbereiche und Forschungsperspektiven

Der Einsatz von KI-Systemen gewinnt sowohl in der Berufswelt als auch in der Berufsbildung rasant an Bedeutung. In der Folge sind (angehende) Lehrkräfte gefordert, sich tiefergehend mit KI sowie deren Anwendung bei der Gestaltung von Lehr-Lern-Prozessen auseinanderzusetzen. Der Beitrag fokussiert auf Basis der Grundlagen und Funktionsweisen von KI die Potentiale von KI als Assistenz für Lehrkräfte bei der Unterrichtsgestaltung. Es wird der Frage nachgegangen, wie eine KI-basierte Lernplattform zur Unterstützung der Unterrichtsplanungskompetenz konzipiert werden kann. Vorgestellt wird dabei das KI-Projekt EDDA, das für die Analyse von Unterrichtsentwürfen und -materialien im Studium, im Vorbereitungsdienst und in der schulischen Berufstätigkeit eingesetzt werden kann. EDDA stellt Rückmeldungen zur Umsetzung ausgewählter didaktischer Merkmale der Unterrichtsplanung bereit und liefert Hinweise zur Reflexion und Weiterentwicklung der Entwürfe. Zudem bietet EDDA eine Infrastruktur für weitere Forschungen zum Lehren und Lernen mit KI.

Schlagwörter: Künstliche Intelligenz – Professionalisierung – Unterrichtsplanung – Wirtschaftspädagogik

Electronic didactic assistance (EDDA) for lesson planning based on artificial intelligence: functionality, areas of application and research perspectives

The use of AI systems is rapidly gaining importance both in the professional world and in vocational education and training. As a result, (prospective) teachers are required to take a closer look at AI and its application for teaching and learning processes. This article focuses on the potential of AI to assist teachers in designing lessons based on the fundamentals and functions of AI. We explore the question of how an AI-based learning platform can be designed to help develop lesson planning skills. The AI platform EDDA will be presented, which can be used to analyze lesson plans and materials during studies, in the preparatory service and in teaching practice. EDDA provides feedback on the implementation of selected didactic criteria of lesson planning and provides information for reflection and further development of lesson plans. EDDA also provides an infrastructure for further research into teaching and learning with AI.

Keywords: artificial intelligence – business education – lesson planning – teacher professionalization

1 Einleitung

Künstliche Intelligenz (KI) ist die Technologie, die in der Berufswelt und im Bildungssystem am meisten diskutiert wird. Während Schlüsselaspekte der heutigen Modelle, wie die Backpropagation, schon in den 1990er Jahren bekannt waren, sind erst durch Fortschritte bei Hardware und verfügbaren Trainingsdaten die Voraussetzungen geschaffen, die Technologie flächendeckend und mit deutlichen Nutzungsvorteilen einzusetzen (Chollet, Kalinowski & Allaire, 2022, S. 20ff.). Wurde KI anfangs nur zur Übernahme von Routinetätigkeiten verwendet, ist sie heute auch in komplexen Tätigkeiten einsatzfähig (Davenport & Kirby, 2016). Die KI als Partnerin unterstützt nun im Sinne der Augmentation, d. h. der Erweiterung des Anwendungspotentials, den Menschen bei (zeit-)aufwendigen Routine- und Nicht-Routinetätigkeiten, um die Ergebnisqualität zu steigern und den Menschen zu entlasten (Jarrahi, Askay, Eshragi & Smith, 2023; Seufert, Guggemos & Ifenthaler, 2021).

Für (angehende) Lehrkräfte sind diese Aspekte vor allem in Bezug auf ihre Kernaufgabe, das Unterrichten (KMK, 2019), von Bedeutung. Dabei ist die Planung des Unterrichts ein hochkomplexer, wenig routinefähiger Tätigkeitsbereich. Im von Stender, Brückmann und Neumann (2015) entwickelten Transformationsmodell der Unterrichtsplanung kommt der Auseinandersetzung mit der Unterrichtsplanung eine besondere Bedeutung zu, indem es die professionelle Kompetenz von Lehrkräften (basierend auf dem Modell von Baumert und Kunter (2006) in konkrete Handlungsskripte übersetzt und damit einen Beitrag leistet, dass professionelle Wissen von Lehrkräften in reale Handlungen zu überführen. Mit anderen Worten: In den Unterrichtsplanungen kombinieren Lehrende Elemente aus ihrer gesamten professionellen Kompetenz, um sich dadurch auf hochwertige Handlungen in komplexen und konkreten Unterrichtssituationen vorzubereiten. In der Folge geht das Modell davon aus, dass sich eine ‚gute‘ Vorbereitung in einer entsprechenden Unterrichtsqualität äußert (Stender et al., 2015, S. 127f.). Diese sollte sich letztendlich in einem hohen Lernerfolg der Lernenden widerspiegeln.

Vor dem Hintergrund dieses Modells ist es daher nicht verwunderlich, wenn zur Förderung der professionellen Kompetenz von Lehrkräften Unterrichtsentwürfe i. d. R. das zentrale Mittel darstellen (Söll & Klusmeyer, 2018; Reinisch, 2021), mit dem auch Innovationen in den Unterricht integriert werden können (Aprea, 2018). Die Fähigkeit zum professionellen Unterrichten entwickelt sich somit aus der kontinuierlichen, begleiteten Planung von Unterricht. Gleichzeitig stellt die Nutzung von detaillierten Unterrichtsentwürfen Lehrkräftebildner*innen vor die Herausforderung, den Erstelnden ausführliches, individuelles Feedback zu geben und eine begleitete Implementation zu ermöglichen. Es gibt somit verschiedene Ansatzpunkte für einen lohnenswerten Einsatz von KI: einerseits in den konkreten Planungsaktivitäten, andererseits in der Begleitung dieser. Hier könnte eine KI-Assistenz ihr Potential

entfalten, komplexe Arbeitstätigkeiten zu erleichtern (Seufert et al., 2021). Auch wenn erste Befunde zum Einsatz von Apps zur Planung von Unterricht (Schrader & Schöb, 2016) oder zu Chatbots im Bildungsbereich (Wollny et al., 2021) vorliegen, ist die Studienlage zur Interaktion von (angehenden) Lehrkräften mit KI-Systemen hingegen noch nicht umfangreich (Jeon & Lee, 2023; Salas-Pilco, Xiao & Hu, 2022).

Um die Möglichkeiten zur Ausgestaltung der Beziehung von KI und Lehrkräften zu erforschen, werden Anwendungen notwendig, die den Lehrkräften eine spezialisierte KI-basierte Assistenz für die Interaktion zur Verfügung stellen. Dieser Beitrag greift diesen Bedarf auf und beschreibt die KI-Plattform EDDA. Sie ist dafür konzipiert, (angehende) Lehrkräfte in ihren Planungsaktivitäten von Unterricht unter Beachtung zentraler didaktischer Prinzipien beruflichen Lehrens und Lernens sowie allgemeiner lernpsychologischer Grundlagen zu unterstützen. EDDA bewertet den Erfüllungsgrad didaktischer Merkmale in Unterrichtsentwürfen und dem dazugehörigen Material auf Basis von 57 Kriterien, prüft die Stimmigkeit zwischen Unterrichtsplanung und -material und unterbreitet individuelle Überarbeitungsvorschläge. EDDA kann daher auf drei Arten genutzt werden:

- ▶ Sie unterstützt Lehr-Lern-Prozesse in allen Phasen der Lehrkräftebildung.
- ▶ Sie verbessert Unterrichtsqualität durch wissenschaftlich fundierte Rückmeldungen.
- ▶ Sie liefert Erkenntnisse über Unterricht durch Subsumtion von Entwürfen und Materialien in wissenschaftliche Modelle.

Ziel des Beitrags ist, die technische Konzipierung von EDDA als Muster für die Entwicklung weiterer Software-Anwendungen vorzustellen, die erwarteten Lehr-Lern-Prozesse bei (angehenden) Lehrkräften zu begründen und Forschungsmöglichkeiten mit der Software offenzulegen. Der Fokus liegt daher nicht auf der Beforschung der Mensch-Maschine-Interaktion an sich, sondern auf der theoretischen Begründung, den Designaspekten der Plattform und auf den Möglichkeiten zur Zusammenarbeit (Augmentation) der KI EDDA als Partnerin. Hierfür wird zunächst hergeleitet, welchen Ansprüchen KI-Systeme gerecht werden müssen, um als Assistenz für Lehrkräfte einsetzbar zu sein (Kapitel 2). In Kapitel 3 wird der konkrete Anwendungsfall am Beispiel von EDDA illustriert. In Kapitel 4 werden dann die Potentiale von EDDA für Forschung und Praxis diskutiert.

2 KI als Assistenz für Lehrkräfte

2.1 KI und die Verarbeitung von natürlicher Sprache

Von grundlegender Bedeutung bei der Erforschung von KI ist der Intelligenzbegriff, den Götz et al. „als Erkenntnisvermögen, als Urteilsfähigkeit, als das Erfassen von Möglichkeiten, aber auch als das Vermögen, Zusammenhänge zu begreifen und

Einsichten zu gewinnen“ (2021, S. 6) bezeichnen. In der KI-Forschung erfolgt die Übertragung genau dieses Problemlösungsansatzes auf technische Systeme. Hierzu definiert Rich (1983): „Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.“ Aktuell bewegen sich die Forschungs- und Entwicklungsaktivitäten auf der Ebene der schwachen KI, auch wenn jüngst mit ChatGPT der bislang größte Schritt in Richtung einer allgemeinen KI gegangen wurde (Buxmann & Schmidt, 2021). Die bisher verfügbaren Systeme sind allerdings noch immer anwendungsspezifisch und erreichen damit nicht die Klassifikationsstufe einer allgemeinen KI (Latif et al., 2023).

Wesentliches Element der Leistung einer KI ist ihre Fähigkeit zum maschinellen Lernen (Buxmann & Schmidt, 2021). Damit ist gemeint, dass Computer ein Problem lösen, indem sie das dafür notwendige Programm selber (weiter)entwickeln, d. h. den dafür notwendigen Algorithmus eigenständig erzeugen (Alpaydin, 2019; Lanquillon, 2019, S. 91). Abb. 1 zeigt drei Typen des eigenständigen Lernens, die i. d. R. differenziert werden: das überwachte, das unüberwachte sowie das verstärkende maschinelle Lernen (Kleesiek, Murray, Strack, Kaissis & Braren, 2020). Während im überwachten Lernen Trainingsdaten und Outputs eingegeben werden müssen (d. h. Muster werden vorgegeben), können Systeme, die auf unüberwachtem maschinellen Lernen basieren, selbstständig Muster in den Trainingsdaten erkennen. Algorithmen des verstärkenden maschinellen Lernens benötigen keinen Trainingsdatensatz, das System lernt durch die Interaktion mit der Umwelt und generiert die notwendigen Trainingsdaten eigenständig (Russel & Norvig, 2021; Welsch, Eitle & Buxmann, 2018).

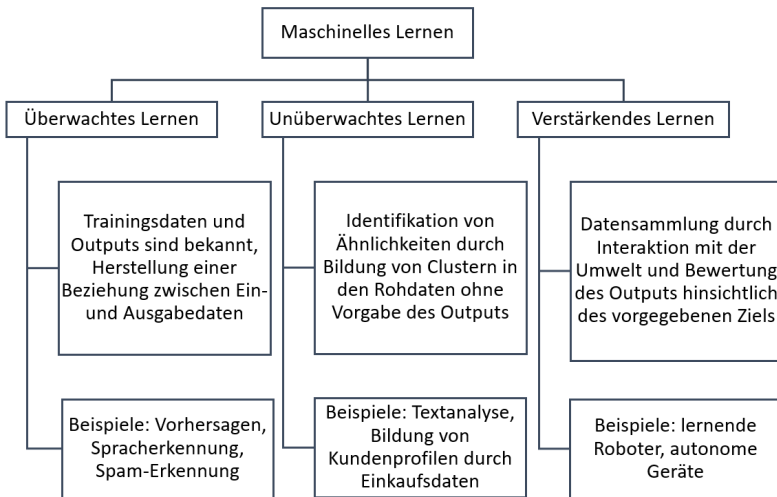


Abbildung 1: Arten maschinellen Lernens

Von den verschiedenen Datenarten, die für ein Training nach dem überwachten bzw. unüberwachten Lernen zur Verfügung stehen, eignen sich Unterrichtsentwürfe in der Lehrkräfteausbildung in besonderer Weise, da diese häufig ein Bestandteil der Bildungsgänge sind. So werden in formellen Bildungssettings Unterrichtsentwürfe in Fließtextform mit Tabellen des geplanten Unterrichtsverlaufs von (angehenden) Lehrkräften angefertigt und für Beratungs- oder Bewertungszwecke genutzt, Unterrichtsmaterial nutzt schriftlich fixierte Kommunikation, z. B. Lern- und Arbeitsaufgaben, Arbeitsblätter oder Auszüge aus Schulbüchern. Auch handlungsorientierte Methoden, wie die Leittextmethode oder die Portfolioarbeit sind auf schriftliche Kommunikation angewiesen.

Zur KI-gestützten Verarbeitung solcher schriftlichen Daten wird das Natural Language Processing (NLP) benötigt. NLP ist ein Sammelbegriff für computergestützte Verfahren, die eine Analyse und Repräsentation der menschlichen Sprache erlauben (Chowdhary, 2020, S. 604; Lane, Howard & Hapke, 2019, S. 4). Aktuell stellen Transformermodelle die zentrale Architektur im NLP-Bereich dar (Tunstall, von Werra, Wolf & Géron, 2022, S. 15). Die Leistungsfähigkeit dieser Modelle resultiert aus dem Self-Attention-Mechanismus (Vaswani et al., 2017). Dieser Mechanismus erlaubt es einer KI, Sprache in ihrem Kontext zu verstehen (Chollet et al., 2022,

S. 367f.). Zudem lassen sich diese Modelle nach dem Prinzip des unüberwachten maschinellen Lernens trainieren (Tunstall et al., 2022, S. 1), das deutlich weniger Ressourcen für die Erstellung von Trainingsdaten erfordert als das überwachte maschinelle Lernen.

Obwohl sich mittlerweile eine große Vielzahl an Transformermodellen finden lässt, gibt es im Wesentlichen zwei ‚Grundarchitekturen‘, die GPT-Modelle (Generative Pretrained Transformer) und die BERT-Modelle (Bidirectional Encoder Representations from Transformers) (Tunstall et al., 2022, S. 1). Die GPT-Modelle, z. B. das neuste Modell GPT-4 von OpenAI (2023), werden i. d. R. nach der Technik des Casual Language Modeling (CLM) trainiert (Tunstall et al., 2022, S. 126, S. 323). Dabei ist es die Aufgabe des Modells, einen vorgegebenen Satzanfang fortzuführen, indem es das nächste Wort vorhersagt. Diese Technik eignet sich vor allem für Aufgaben, die Textgenerierung beinhalten.

Die BERT-Modelle hingegen nutzen i. d. R. das Masked Language Modeling (MLM). Hierbei muss das Modell ähnlich wie bei einem Lückentext das fehlende Wort vorhersagen, das durch [MASK] gekennzeichnet ist (Tunstall et al., 2022, S. 9). Im Unterschied zum CLM, wo nur die Wörter links im Satz relevant sind, nutzt MLM sowohl die Wörter links als auch rechts von der Lücke. Im Gegensatz zu CLM erlernt das Modell damit in besonderer Weise, den gesamten Kontext in einem Text nutzbar zu machen. Allerdings befähigt diese Methode nicht so gut dazu, Texte zu generieren wie das CLM, da diese Technik nicht direkt auf die Generierung von Texten abzielt (Abb. 2).

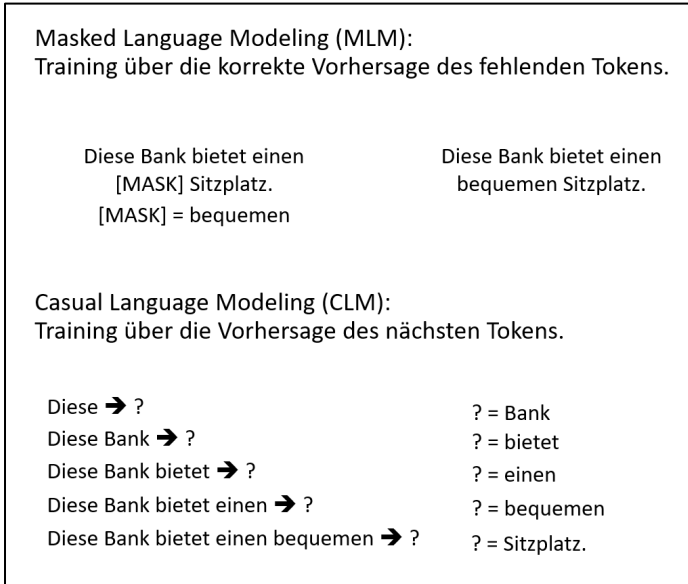


Abbildung 2: Vergleich der elementaren Trainingstechniken für Transformermodule

CLM und MLM dienen allerdings nur dazu, das sprachliche Verständnis einer KI zu trainieren. Konkrete Anwendungen sind damit noch nicht verbunden. Hierfür hat sich ein mehrstufiges Vorgehen etabliert (Abb. 3), bei dem ein Transformer zunächst mit einem großen Korpus an Textmaterial (z. B. Wikipedia) trainiert wird (Tunstall et al., 2022, S. 8). Dieses Modell wird dann in einem zweiten Schritt mit einem – im Vergleich zum ersten Schritt – kleinen und spezialisierten Textkorpus an die Domäne des späteren Einsatzes angepasst. Erst im dritten Schritt erfolgt dann das Training für die eigentliche Anwendungsaufgabe, indem das Transformermodell um weitere Schichten ergänzt wird, z. B. für die Klassifikation von Text (Tunstall et al., 2022, S. 8).

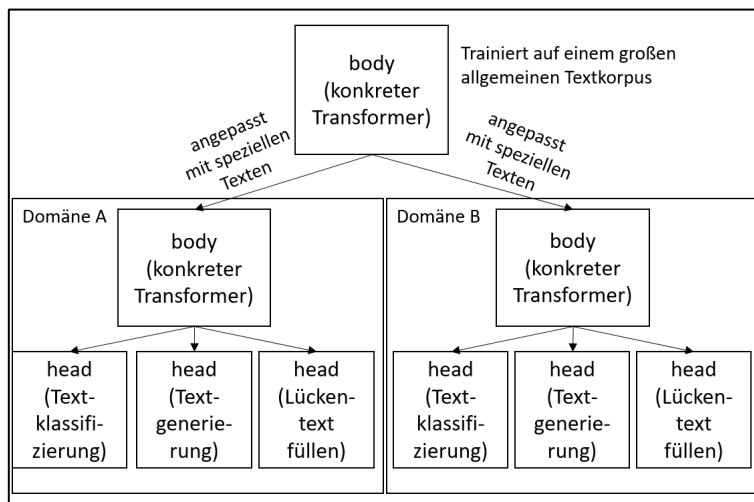


Abbildung 3: Anwendung von Transformermodellen

Technisch bedeutet dies, dass das Transformermodell den sog. body bildet, auf den je nach Anwendungsfall ein anderer head platziert wird (Tunstall et al., 2022, S. 6). Der Grund für diese Vorgehensweise liegt darin, dass dadurch gute Ergebnisse mit einem deutlich reduzierten Rechen- und Datenaufwand erzielt werden können. Dies kann den Energiebedarf und den CO₂-Verbrauch der Systeme reduzieren.

2.2 KI und professionelles Handeln von Lehrkräften

Über die Transformermodelle werden neue Arten der Zusammenarbeit von Anwender*innen mit KI-Systemen möglich. Seufert et al. (2021) leiten ab, dass nicht mehr die Verdrängung menschlicher Tätigkeiten durch diese im Fokus steht, sondern die Zusammenarbeit zwischen ihnen. Damit wird das Verhältnis partnerschaftlich, so dass jeweils die Stärken des Menschen respektive der Maschine genutzt werden (Davenport & Kirby, 2016; Wesche & Sonderegger, 2019). Ziel der Interaktion zwischen KI und (angehender) Lehrkraft muss es daher sein, die Lehrkraft in ihrer Professionalisierung und ihrem professionellen Handeln zu unterstützen.

Während das Transformationsmodell der Unterrichtsplanung (Stender et al., 2015) beschreibt, dass Unterrichtsplanungen dazu beitragen, die professionelle Kompetenz von Lehrkräften in Handlungsskripte zu überführen, um damit die Wahrscheinlichkeit hochwertiger, realer Handlungen zu erhöhen, gibt es wenig Auskunft darüber, wie die Struktur dieser Handlungen zu modellieren ist. Diese Lücke kann der strukturanalytische Ansatz von Oevermann (1996) schließen (Helsper, 2011, S. 149).

In diesem Ansatz zeichnet sich professionelles Handeln dadurch aus, dass Wissen für eine konkrete Situation aus zwei miteinander verbundenen Perspektiven genutzt wird: zum einen interventionspraktisch durch Rekonstruktion des individuellen Einzelfalls in der Art, dass der Einzelfall durch Deutung, Interpretation und Auslegung in standardisiertes Wissen subsumiert werden kann. Zum anderen ingenieurl, d. h. durch Deduktion bzw. Subsumtion von Konsequenzen aus standardisiertem Wissen für den Einzelfall (Oevermann, 2008, S. 58f.). Abb. 4 verdeutlicht den Prozess.

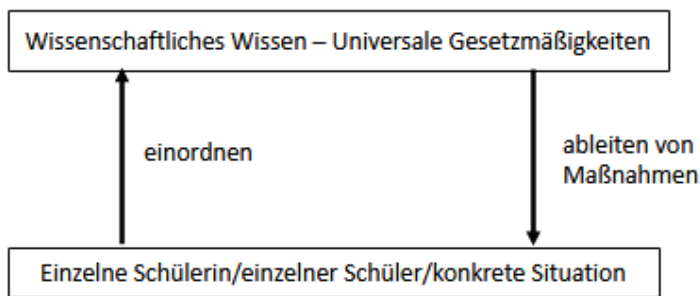


Abbildung 4: Wissensnutzung im strukturanalytischen Ansatz von Oevermann (1996)

Um über die Analyse von Unterrichtsentwürfen und dem dazugehörigen Material einen Beitrag zur Professionalisierung von (angehenden) Lehrkräften zu leisten, wird die geplante Unterrichtssituation mit Hilfe der KI in wissenschaftliche Kategorien eingeordnet, aus denen Empfehlungen zur Verbesserung des konkreten Unterrichts gemeinsam mit der Lehrkraft abgeleitet bzw. entwickelt werden. Eine solche KI würde somit dem Ziel Rechnung tragen, Unterrichtspraxis und Wissenschaft stärker zu verzahnen.

2.3 Schlussfolgerungen für die Entwicklung von KI als Assistenzsysteme in der Lehrkräftebildung

Die hier dargestellten Forschungsstände bieten Implikationen für die Entwicklung einer KI im Rahmen der Lehrkräftebildung. Die Modellierung professionellen Handelns nach Oevermann (1996) und die Forderung in Learning Analytics, Anwendungen an wissenschaftlichen Theorien zu orientieren (Wong et al., 2019, S. 4), legen nahe, dass eine KI-basierte Assistenz wissenschaftliche Konzepte für die Bewertung

von Unterrichtsentwürfen nutzen sollte. Gemäß Oevermanns Theorie (2008) muss die KI in diesem Kontext folgende Schritte vollziehen: Unterrichtsentwürfe in wissenschaftliche Kategorien einordnen, Konsequenzen aus wissenschaftlichen Modellen ableiten und diese auf das Material und die Planung beziehen (Abb. 5). Nur wenn eine KI die Struktur professionellen Handelns annäherungsweise erlernt, kann sie Lehrkräfte beim professionellen Handeln bzw. den dazugehörigen Professionalisierungsprozessen unterstützen.

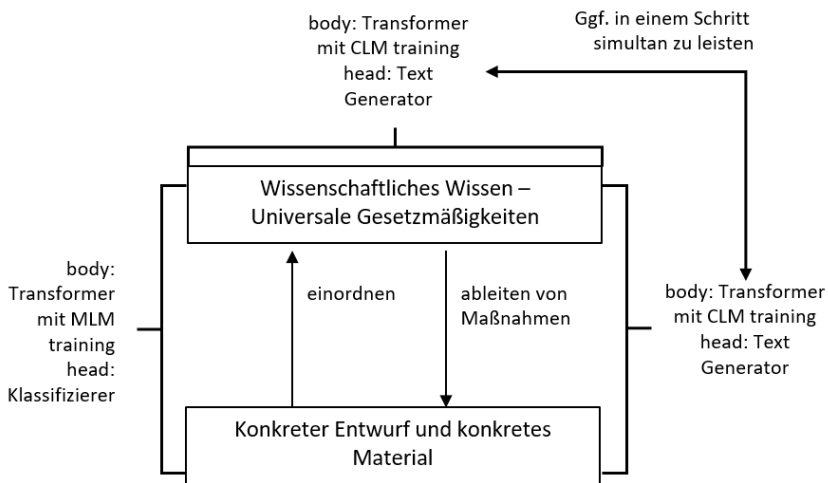


Abbildung 5: Ansatzpunkte für KI-Modelle im professionellen Handeln

Anmerkungen: MLM = Masked Language Modeling, CLM = Casual Language Modeling (siehe ausführlich Abschnitt 2.1)

Die Prozesse der Subsumtion und der Ableitung müssen nun technisch umgesetzt werden:

1. Subsumtion von konkreten Fällen: Für das Training der KI im Rahmen der Subsumtion erscheint das überwachte maschinelle Lernen am besten geeignet (Abschnitt 2.1), denn es stellt eine Beziehung zwischen den Trainingsdaten (hier in Form von Unterrichtsentwürfen und Materialien) und dem Output (hier wissenschaftliche Kategorien) her (Lanquillon, 2019, S. 96f.). Dies ermöglicht der KI, gezielt wissenschaftliche Begriffe und Modelle anzuwenden. Die Bewertungen sind somit mit dem aktuellen Forschungsstand kompatibel und unterstützen die Professionalisierung der Lehrkräfte. Unüberwachtes oder verstärkendes maschinelles Lernen gewährleisten diese Anpassungsfähigkeit nicht. Hier entwickelt die KI eigene Begriffe

und Strukturen, die nicht unbedingt dem Forschungsstand entsprechen. In der Folge muss die KI (genauer der head des Transformers, siehe Abschnitt 2.1 und Abb. 3) eine Klassifikations- bzw. Regressionsaufgabe lösen (Lanquillon, 2019, S. 96). Zudem ermöglichen Klassifikatoren und Regressionsanalysen die Bewertung der Lernprozessqualität mit verschiedenen Metriken. Dazu gehören Kennzahlen aus der qualitativen Inhaltsanalyse wie Krippendorffs Alpha (Krippendorff, 2019), Gwets Übereinstimmungskoeffizient (Gwet, 2014) oder dem Iota-Konzept (Berding & Pargmann, 2022). Diese können beschreiben, wie gut die KI menschliche Einschätzungen replizieren kann.

Dieser Ansatz ist gleichwohl mit Herausforderungen verbunden:

- ▶ Wenige Daten: Unterrichtsentwürfe enthalten z. B. Aussagen über einzelne Lernende oder Schulen. Diese Daten gilt es zu schützen, sodass die Generierung von Datenmaterial aufwendig ist und zu kleinen Datensätzen führt.
- ▶ Unausgewogene Daten: Idealerweise erfolgt das Training von Klassifikatoren mit ausgewogenen Daten, in denen die Häufigkeiten aller Kategorien etwa gleich sind. Unausgewogene Daten führen dazu, dass der Klassifikator sich auf häufige Kategorien fokussiert, um einen geringen globalen Fehler zu erzielen. Dies hat zur Folge, dass seltene Kategorien schnell ‚übersehen‘ werden (Haixiang et al., 2017, S. 220f.). Wie die Analyse von didaktischem Material und Schulbuchanalysen von Bloemen (2011), Stütz, Berding, Reincke und Scheper (2022) oder Riebenbauer (2022) zeigen, sind unausgewogene Daten häufig anzutreffen.
- ▶ Komplexe Daten: Zur Beschreibung von Lehr-Lern-Prozessen hat die Wissenschaft immer bessere Modelle entwickelt, die sich in entsprechend komplexen Konstrukten manifestieren.
- ▶ Aufwendige Datengenerierung: Wenn Daten vorliegen, müssen diese für das überwachte maschinelle Lernen zunächst durch Menschen (z. B. in Form einer qualitativen Inhaltsanalyse) bewertet werden, um die Zieldaten für das Training zu generieren.

Für das zu nutzende Sprachmodell, den Transformer, sind vor allem Ansätze nach MLM zielführend, da sie den gesamten Text kontextsensitiv nutzen können. Aktuell empfehlen sich besonders Modelle aus der BERT-Familie. Generative Modelle wie GPT, die mit CLM trainiert werden, sind aufgrund ihrer geringeren Kontextsensitivität weniger geeignet.

2. Ableitung von Handlungsempfehlungen: Anders sieht es aus, wenn eine erfolgreiche Klassifizierung erfolgt ist und eine KI konkrete Verbesserungsvorschläge generiert. Das würde die Nutzung von GPT-Modellen mit CLS-Training empfehlenswert machen. Das gezielte und wissenschaftlich überprüfbare Training dieses Assistenzteils ist herausfordernd, da eine große Anzahl von Materialbewertungen und

dazugehörigen präzisen und ausformulierten Empfehlungen erforderlich ist. Vor dem Hintergrund dieser Herausforderungen sowie der dargelegten technischen Konzepte erfolgte die Entwicklung von EDDA.

3 EDDA als Anwendung- und Forschungsobjekt für eine KI in der Lehrkräftebildung

3.1 Darlegung der Planungsschritte und Entwicklung der Analysekriterien

Zur Unterrichtsplanung existieren zahlreiche neuere und ältere Modelle (z. B. das Berliner Modell oder das Hamburger Modell). Obgleich die zugrundeliegenden Paradigmen variieren, haben sie Überlegungen zur Lerngruppe, zum Inhalt, zum Verlauf des Unterrichts, zu Lern- und Kompetenzzielen oder zum Medien- und Methodeneinsatz gemeinsam (Schmaltz, 2019). Ausgegangen wird von inhaltlichen und lerngruppenspezifischen Voraussetzungen. So folgt der moderne Planungsprozess von Unterricht in gewissem Maße vereinheitlichten Schritten, für die das Vokabular zwar variiert, deren didaktischen Begründungen aber im Kern Ähnlichkeiten aufweisen. Die Forschungsgruppe hat diese Überlegungen in Abb. 6 als Prozess zusammengeführt, der Rückkopplungen zwischen den einzelnen Planungsschritten beinhaltet. Das Durchlaufen dieser Planungsschritte dient als Ausgangspunkt der Arbeit mit EDDA.

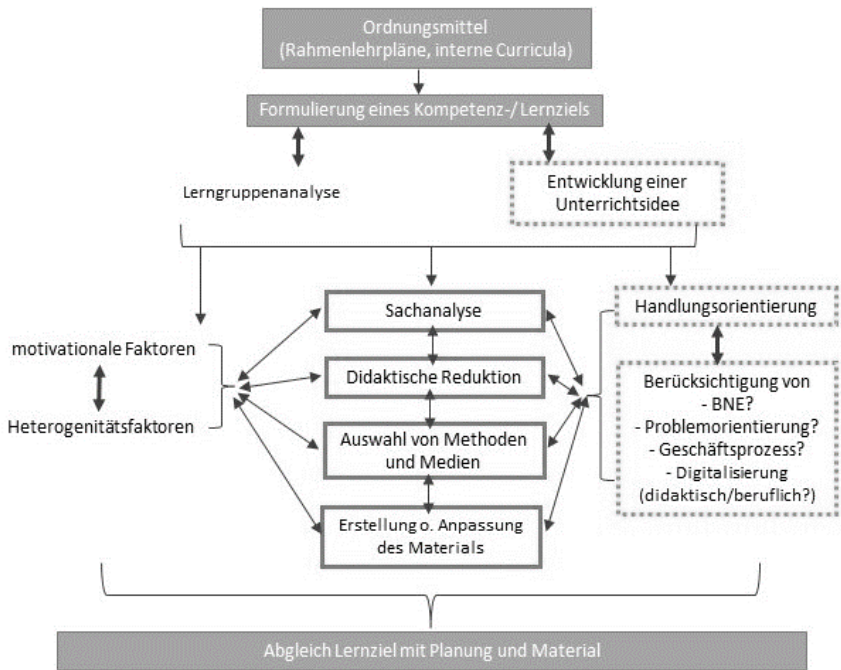


Abbildung 6: Unterrichtsplanungsprozess

Anmerkungen: BNE = Bildung für nachhaltige Entwicklung'

Aus der Vielfalt der didaktischen Modelle und Planungsdimensionen für Unterricht wurden allgemein- und fachdidaktische Merkmale ausgewählt. Die allgemeinen Dimensionen beinhalten die Grundlagen des Planens und die Berücksichtigung individueller motivationaler Tendenzen zur Steigerung des Lernerfolgs. Um der Didaktik der beruflichen Bildung gerecht zu werden, wurden ferner aus der Handreichung zur Erarbeitung von Rahmenlehrplänen und den Präambeln der Rahmenlehrpläne traditionelle Analyse Kriterien wie die Handlungs- und Geschäftsprozessorientierung und die Problemorientierung ausgewählt. Außerdem wurden mittels der Standardberufsbildpositionen Nachhaltigkeit und Digitalisierung (BIBB, 2021) als moderne Perspektiven integriert. Aus der Vielfalt didaktischer Modelle und Planungsdimensionen von Unterricht wurden somit einige zentrale herausgegriffen. Hervorzuheben ist, dass die Empfehlungen von EDDA stark von der Operationalisierung dieser Dimensionen abhängen und somit nicht als allgemeingültig betrachtet werden sollten. Im Zuge einer anders gelagerten Operationalisierung und Ausrichtung der Empfeh-

lungen könnten sich bei anderen Definitionen auch andere Implikationen für die hochgeladenen Entwürfe ergeben. In Anschluss an Oevermann (1996) nutzt EDDA nun jedoch die folgenden 57 Analyse Kriterien, um die Unterrichtsentwürfe und -materialien der Nutzenden zu kategorisieren und entsprechende Empfehlungen auszusprechen (Abbildung 7):

Basisdimensionen 3 Merkmale	Problemorientierung 1 Merkmal	Motivationsförderung 5 Merkmale
Authentizität der Handlungen und Operationen 4 Merkmale	Vollständigkeit der Handlungen 6 Merkmale	Prozessorientierung 4 Merkmale
Sustainability Development Goals 17 Merkmale	Integration von Nachhaltigkeit 7 Merkmale	57 Merkmale insgesamt
Digitale Lehr-Lern-Technologien 4 Merkmale	Digitale Technologien der Berufswelt 6 Merkmale	

Abbildung 7: Übersicht über die Kategorien und Merkmale

3.1.1 Basisdimensionen Strukturierung, Aktivierung und Vernetzung

Die Unterrichtsplanungsprozesse lassen sich in die drei Basisdimensionen Strukturierung, Aktivierung und Vernetzung differenzieren (Riebenbauer, 2022). Die Dimension der Strukturierung beschreibt die Phasierung sowie die Sequenzierung des Unterrichts und die Begründung der Planungsschritte. Die Aktivierung bezieht sich auf die Lernenden, die durch praxis- und problemorientierte Situationen selbstständig anwenden und üben können. Bei der Vernetzung geht es primär um die Herstellung von Zusammenhängen mit anderen Lernfeldern bzw. -bereichen oder innerhalb desselben Gebiets.

3.1.2 Problemorientierung

Dörner (1979, S. 10) zufolge liegt dann ein Problem vor, wenn ein unerwünschter innerer oder äußerer Zustand vorliegt, der mit zur Verfügung stehenden Mitteln nicht beseitigt werden kann. Entlang der Theorie der kognitiven Entwicklung von Piaget (1977) kann ein unerwünschter Zustand dadurch beseitigt werden, dass neue kognitive Schemata entwickelt oder bestehende erweitert werden.

3.1.3 Motivationsförderung

Die Selbstbestimmungstheorie nach Deci und Ryan (1993) besagt, dass drei Grundbedürfnisse existieren: Autonomieerleben, Kompetenzerleben und soziale Eingebundenheit, um eine hochwertige Motivation hervorzurufen. Bloemen (2011) identifiziert die drei Subkategorien Problem-, Lösungsweg- und Ergebnisoffenheit als notwendig in der Beförderung des Autonomieerlebens.

3.1.4 Prozessorientierung

Ein prozessorientierter Unterricht spricht nicht nur einzelne Tätigkeiten an, sondern setzt sie in Bezug zu vor- und/oder nachgelagerten Aktivitäten oder der gesamten Wertschöpfungskette des Unternehmens (Bloemen, 2011; Stütz et al., 2022).

3.1.5 Handlungsorientierung und Authentizität

Der Ansatz der Situated Cognition (Brown et al., 1989) fordert authentische Aktivitäten, d. h. die Aktivitäten ähneln jenen, die dem späteren Anwendungskontext der Lernenden entsprechen. In dieser Kategorie wird differenziert, ob Lernende die Operationen im Sinne Aebli (2011) gedanklich durchdringen oder auch praktisch durchführen müssen. Weitere Analysekriterien der Handlungsorientierung sind die Phasen einer vollständigen Handlung: Informieren, Planen, Entscheiden, Ausführen, Kontrollieren und Auswerten (Hacker & Sachse, 2014).

3.1.6 Integration von Nachhaltigkeit

Die Umsetzung der Nachhaltigkeit in den Unterrichtsentwürfen wird mittels zweier Skalen betrachtet. Dies sind die 17 Sustainable Development Goals der Vereinten Nationen und die Beförderung von Lernenden zu Sustainable Change Agents (Berding, Slopinski & Frerichs, 2020; United Nations, 2015).

3.1.7 Umsetzung von inhaltlicher und methodischer Digitalisierung

Meta- und Meta-Meta-Studien liefern Hinweise darauf, wie digitale Technologien methodisch im Unterricht eingesetzt werden sollten, um lernförderlich zu sein. Dies berührt die Bereiche des Technologieeinsatzes (Bernard et al., 2009), der digitalen Zusammenarbeit (Tamim, Bernard, Borokhovski, Abrami & Schmid, 2011), des Feedbacks (van der Kleij, Feskens & Eggen, 2015) und der Art des Einsatzes (Schmid et al., 2009). Die inhaltliche Digitalisierung orientiert sich an den Kriterien von Krämer, Jordanski und Goertz (2017) und unterscheidet den Einsatz allgemeiner und spezieller Technologien der Berufswelt, die Verbalisierung von Informationen aus diesen Technologien, die Nutzung von Technologien zur eigenständigen Aneignung von Fähigkeiten, die Beachtung von Rahmenbedingungen des Technologieeinsatzes sowie die Gestaltung von Medienanwendungen der Berufswelt.

Nachfolgend wird erläutert, wie diese Kategorien technisch in EDDA integriert wurden.

3.2 Technische Umsetzung

Die Forschungsgruppe hat zum Zweck der technischen Umsetzung das Software-Package AI for Education entwickelt (Berding, Pargmann, Riebenbauer, Rebmann & Slopinski, 2023). Es steht aktuell in der Version 0.3.0 über CRAN zur Verfügung und zielt drauf ab, Forschenden ohne große Programmierkenntnisse die Entwicklung und Anwendung von KI unter den speziellen Datenstrukturen der Erziehungswissenschaften zu ermöglichen. Es wird stetig weiterentwickelt.

Für EDDA wird aktuell das bereits vortrainierte Modell German BERT base von Chan et al. (2020) verwendet, welches frei über HuggingFace bezogen werden kann. Das Modell wurde mit allgemeinem Datenmaterial trainiert (z. B. Wikipedia) und wird gegenwärtig in einer noch nicht auf die BWP adaptierten Variante eingesetzt. Das Training der Klassifizierer basiert auf Daten aus Studien zu Unterrichtsentwürfen und Lehr-Lern-Material, insgesamt 5 983 Trainingsdatensätze. Darin sind aktuell 884 Unterrichtsentwürfe (ohne Material), 420 Materialien aus dem Anhang von Unterrichtsentwürfen sowie 4 679 Schulbuchaufgaben enthalten. Die bislang erzielte Leistungsfähigkeit der KI wird auszugsweise in Tab. 1 dargestellt.

Tabelle 1: Metriken zur Reproduktion menschlicher Einschätzungen durch die KI bei für die KI unbekanntem Datenmaterial am Beispiel der vollständigen Handlung und nachhaltigen Handlungskompetenz

Konstrukt	Dynamischer Iota Index	Krippendorffs Alpha	Prozentuale Übereinstimmung	Gwets AC
Vollständige Handlung				
Informieren	.421	.312	.743	.588
Planen	.596	.098	.813	.762
Entscheiden	.517	.147	.772	.687
Durchführen	.512	.386	.799	.700
Kontrollieren	.443	.519	.764	.538
Evaluieren	.819	.019	.926	.920
Nachhaltige Handlungskompetenz				
Fachwissen	.659	.474	.826	.789
Methodeneinsatz	.649	.631	.831	.787
Systemwissen	.816	.508	.911	.903
Kritische Haltung	.769	.282	.874	.862
Entscheidungsfähigkeit	.878	.196	.926	.922
Reflexivität	.857	.209	.921	.917
Sozialkompetenz	.649	.670	.837	.786

Anmerkungen: Gwets AC = Gwets agreement coefficient

Normalerweise werden Klassifizierer dadurch erstellt, dass dem vortrainierten Transformer (body) ein Klassifikationskopf (head) hinzugefügt und das gesamte Modell auf die Klassifikationsaufgabe trainiert wird. Das Package AI for Education (Berding et al., 2023) und damit auch EDDA trainieren hingegen nur den Klassifikationskopf und lassen den Transformer unverändert. Dadurch müssen deutlich weniger Parameter trainiert werden. Zudem zeichnen sich Unterrichtsentwürfe durch lange Texte aus, sodass diese in mehrere Textabschnitte geteilt werden müssen, da BERT i. d. R. nicht mehr als 512 Token verarbeiten kann (Khurana, Koli, Khatter & Singh, 2023, S. 3735). Dies macht ebenfalls technisch i. d. R. eine Trennung von Transformer und Klassifikationskopf notwendig und ermöglicht gleichzeitig einen ressourcenschonenden Trainingsprozess. Die trainierten Klassifizierer können dann in verschiedene Softwareanwendungen wie z. B. EDDA integriert werden. Abb. 8 verdeutlicht die dazugehörige Systematik.

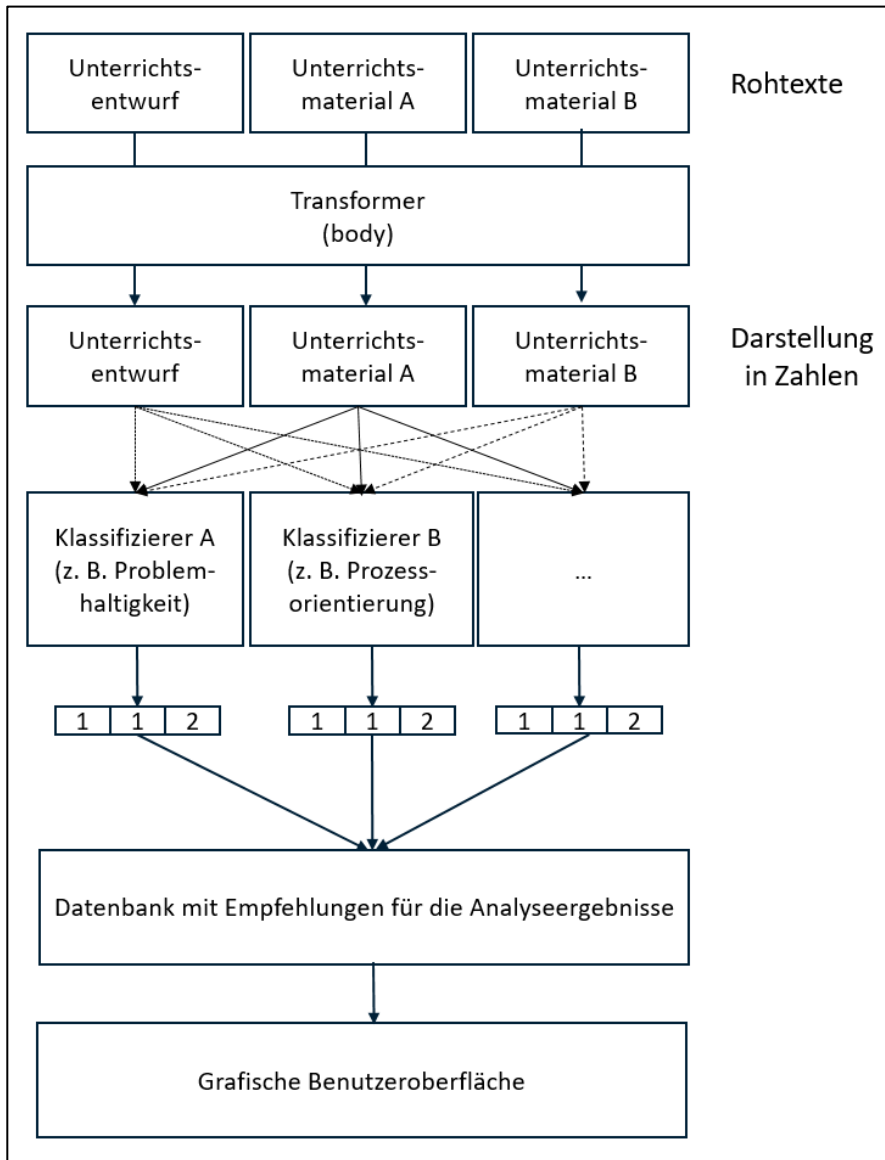


Abbildung 8: Schematische Darstellung des Verarbeitungsprozesses in der Anwendung von EDDA

Nachfolgend wird anhand eines beispielhaften Arbeitsprozesses mit EDDA illustriert, wie diese technischen Aspekte konkret umgesetzt werden. Dabei wird exemplarisch das Kriterium Digitalisierung dargestellt.

3.3 Analyse von Entwürfen und Materialien mit EDDA

Der erste Arbeitsschritt in der Verwendung von EDDA entspricht bei Oevermann (1996) dem Einordnen der konkreten Materialien, d. h. des eingestellten Entwurfs und der dazugehörigen Materialien, in wissenschaftliche Kategorien. Das Grundprinzip zur Nutzung des Tools ist immer gleich: Erst laden die Nutzenden einen Entwurf und ggf. Unterrichtsmaterial im Browser hoch und wählen die gewünschten Dimensionen (Abschnitt 3.1) aus. Zunächst erscheint eine allgemeine Übersichtsseite. Hier werden die erreichten Prozentsätze übersichtlich dargestellt (Abb. 9).

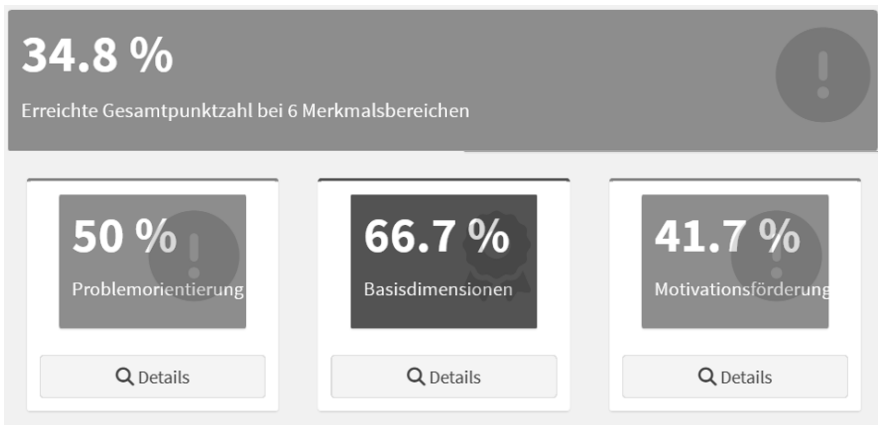


Abbildung 9: Auszüge aus der Übersichtsseite

Nach dem Upload werden die Texte in Zahlensequenzen transformiert, die ihre inhaltliche Bedeutung möglichst gut repräsentieren. Diese Sequenzen stellen den Input für die Klassifizierer dar, die auf dieser Grundlage die Ausprägung der verschiedenen didaktischen Merkmale einschätzen. Es liegen für jedes Dokument (Entwurf und Materialien) zu allen didaktischen Merkmalen eine Einschätzung vor. Zu den geschätzten Werten werden dann passende Empfehlungen aus einer Datenbank geladen und auf der grafischen Benutzeroberfläche für Lehrkräfte dargestellt.

Exemplarische Analyse und Ausgabe:

Durch ‚Details‘-Buttons wird Feedback zu den gewählten Dimensionen angezeigt, wobei die Tiefe variabel ist, abhängig von der Anzahl der ausgewählten Dimensionen. Die Detailseiten folgen einem logischen Aufbau mit einem grafischen und ei-

nem textbasierten Teil. Der grafische Teil umfasst eine Ergebnisanalyse (Abb. 10), eine Kohärenzanalyse und eine Sicherheitsanalyse (Abb. 13).

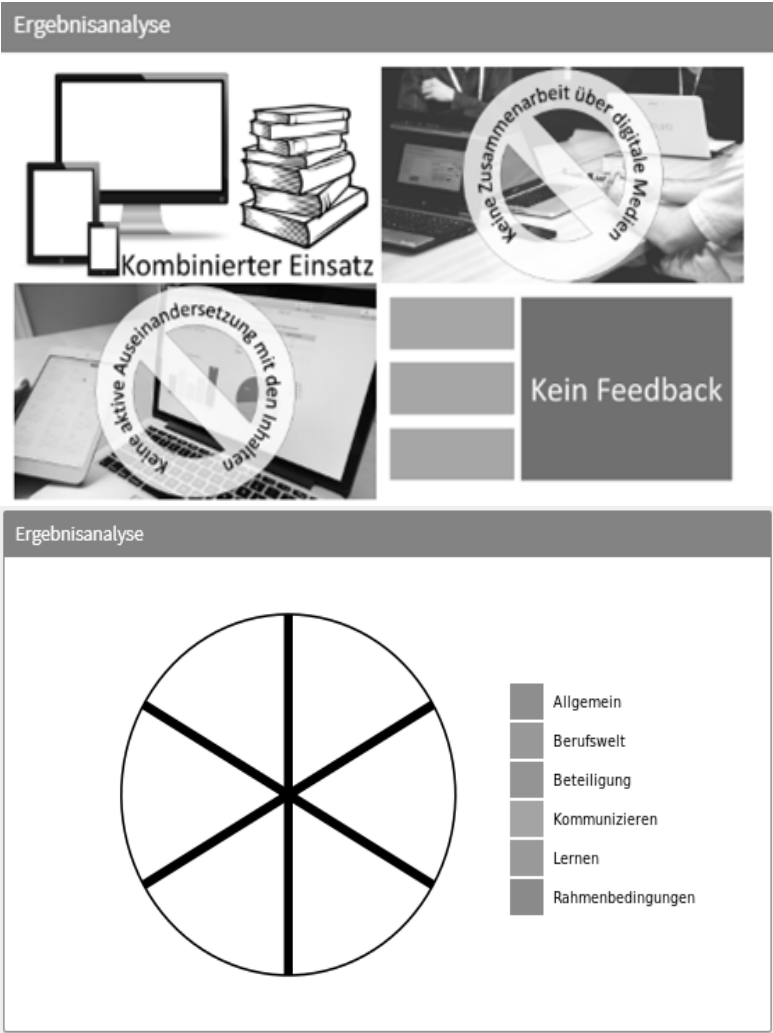


Abbildung 10: Darstellung der Ergebnisanalysen der Skalen ‚Lehr-Lern-Technologien‘ (oben) und ‚Technologien der Berufswelt‘ (unten)

Die Ergebnisanalyse (Abb. 10) stellt EDDAs Einschätzungen graphisch und textbasiert dar. Die Empfehlungen (Abb. 11; Abb. 12) wurden von der Forschungsgruppe manuell entwickelt, um eine möglichst hohe Qualität und Verlässlichkeit der Empfehlungen und die Anschlussfähigkeit an bestehende wissenschaftliche Modelle und Theorie sicherzustellen. Die Grafiken fokussieren auf den Kern der Kategorie, z. B. den Technologieeinsatz. Im Beispielentwurf kombinieren Lernende digitale und analoge Technologien, kooperieren jedoch nicht mittels digitaler Medien und erhalten kein digitales Feedback. Dies tritt häufig auf, wenn traditionelle Methoden durch digitale ersetzt werden, ohne interaktive Elemente zu integrieren. Das Feedback ist nicht immer so ausführlich wie in Abb. 12, wenn z. B. nur wenige Dimensionen ausgewählt sind oder die Kategorien gut erfüllt sind.



Abbildung 11: Beispiele für aufklappbare Textboxen

Technologie zur Zusammenarbeit

Kooperation und Kollaboration

Meine Analysen haben ergeben, dass Sie zwar digitale Technologien in Ihrem Unterricht einsetzen, die Schüler:innen aber weder kooperativ noch kollaborativ mit ihnen arbeiten. Vielleicht bearbeiten die Lernenden Aufgaben in Einzelarbeit im Lernmanagementsystem oder sie arbeiten nicht kollaborativ im selben Dokument, recherchieren an eigenen Endgeräten oder nur Sie als Lehrkraft nutzen digitale Technologien.

▼ Um die Zusammenarbeit der Lernenden via digitaler Technologien zu verbessern, können Sie hier für ein paar methodische Anregungen klicken.

- **Etherpads:** Etherpads sind cloudbasierte Textverarbeitungsprogramme, auf die von verschiedenen Endgeräten zur selben Zeit zugegriffen werden kann. So können verschiedene Lernende zeitgleich an einem Dokument arbeiten oder die Lösungen der anderen Schüler:innen kommentieren, überarbeiten und ergänzen. Einige Lernmanagementsysteme haben Etherpad-Plug-Ins. Ein frei zugängliches Etherpad, gehostet von der Universität Hamburg, finden Sie bspw. hier.
- **TaskCards:** TaskCards sind ein digitales Kollaborationstool, auf welchem a) Informationen unterschiedlich organisiert werden können und b) Lernende Ergebnisse sammeln können. Außerdem können Zeitstrahle oder Lernlandkarten erstellt werden. Auch die Individualisierung in Form eines Wochenplans ist möglich, da verschiedene Bereiche der TaskCard für verschiedene Gruppen freigeschaltet oder verborgen werden können.
- **LearningApps, flinga.fi oder Oncoo:** Diese drei Tools können wie digitale Whiteboards eingesetzt werden, um bspw. die Strukturlegetechnik anzuwenden, Concept Maps und Mind Maps zu erstellen oder Informationen anderweitig grafisch strukturiert aufzubereiten.
- **Cloudbasierte Dokumente:** Anstelle von Textverarbeitungs-, Tabellenkalkulations-, Präsentations- oder Datenbankprogrammen, die nur stationär am digitalen Endgerät bedient werden können, gibt es auch cloudbasierte Programme, die es den Lernenden ermöglichen, parallel an Textdokumenten, Tabellen oder Präsentationen zu arbeiten und das vertraute Layout aufweisen, welches bei Etherpads häufig nicht der Fall ist.

Abbildung 12: Exemplarische Darstellung einer ausgeklappten Textbox

Die Kohärenzanalyse (Abb. 13, links) zeigt Unterschiede zwischen dem erstellten Unterrichtsmaterial und dem Entwurf auf. Die Nutzenden können diese Abweichungen nun kritisch überprüfen und entsprechend das Material im Einklang mit der Zielsetzung des Entwurfs anpassen. Die Sicherheitsanalyse (Abb. 13, links) bewertet die Verlässlichkeit von EDDAs Einschätzungen. In Abb. 13 wird deutlich, dass die Sicherheit in einigen Kategorien weniger ausgeprägt ist als in anderen. Es kann nun kritisch geprüft werden, ob die Aussagen der KI inhaltlich schlüssig sind. Das kann auch Bestandteil von Feedbackgesprächen mit Dozierenden sein.

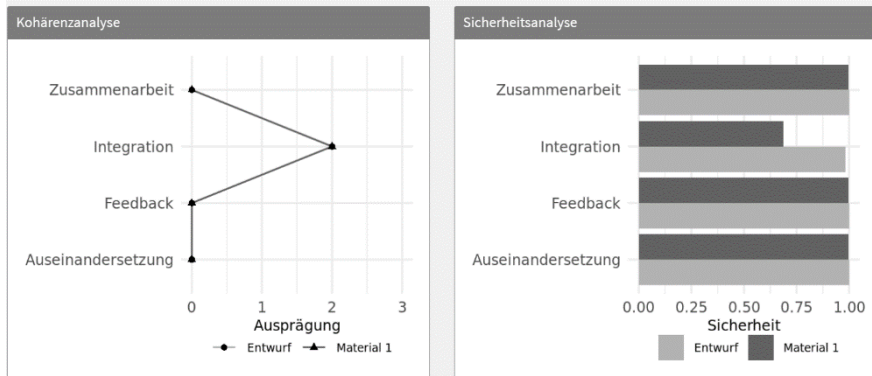


Abbildung 13: Kohärenzanalyse (links) und Sicherheitsanalyse (rechts) zur Skala „Lehr-Lern-Technologien“

4 Diskussion von Anwendungsmöglichkeiten für Praxis und Forschung sowie Ausblick

4.1 Diskussion: Einsatz in der Lehrkräftebildung und der Berufstätigkeit

EDDA wurde mit der Intention konzipiert, in allen Phasen der Lehrkräftebildung einsetzbar zu sein. Im Sinne des Transformationsmodells der Unterrichtsplanung (Stender et al., 2015) setzt die KI-Assistenz an den Unterrichtsentwürfen bzw. Unterrichtsplanungen an, also an dem Katalysator, der einen wesentlichen Beitrag dazu leistet, die professionelle Kompetenz von (angehenden) Lehrkräften in reale Handlungen zu überführen. Damit hat EDDA das Potential, nicht nur die professionelle Kompetenz zu fördern, sondern auch die Chancen auf ihre Anwendung in der Praxis zu steigern.

Durch die Modellierung dieser Handlungen nach dem Ansatz von Oevermann (1996) wird es zudem möglich, wissenschaftliche Modelle und praktische Handlungen in konkreten Situationen miteinander zu verbinden. Hierin liegt ein weiteres Potential der KI-Assistenz, die auf diese Weise zu einer Verzahnung von Wissenschaft und Praxis beiträgt. So bietet EDDA (angehenden) Lehrkräften die Möglichkeit, ihre Unterrichtsentwürfe hochzuladen und evaluieren zu lassen. Dabei unterscheiden sich die Anreize, EDDA zu nutzen, je nach Ausbildungsabschnitt und Berufserfahrung. Übergreifend relevant ist aber die Passung von Unterrichtsplanung und -material. Gerade wenn die Planenden noch wenig Erfahrung haben, wie Materialien zielführend entworfen werden sollten, kann die Analyse der Passung sinnvoll

sein. So bereitet EDDA grafisch auf, in welchen Bereichen Abweichungen bestehen und bietet Reflexionspotential für die Anwender*innen.

Beim Einsatz im Studium sind didaktische Konzepte und Gestaltungskriterien zumindest theoretisch bekannt, die didaktische Transformation in konkreten Unterricht ist aber aufgrund der Vielfalt an Aspekten herausfordernd. Unterrichtseinheiten werden meist in begleitenden Lehrveranstaltungen oder Praxisphasen geplant, in denen erfahrene Lehrkräfte oder Lehrkräftebildner*innen beratend zur Seite stehen. Je nach Größe des Seminars oder der Anzahl der zu betreuenden Praktikant*innen kann dies den Umfang des Feedbacks beeinflussen. EDDA hat beim Einsatz im Studium somit zwei Anwendungsbereiche: (1) in der individuellen Auseinandersetzung mit dem eigenen Unterrichtsentwurf und (2) in der beratenden Auseinandersetzung mit den Entwürfen der Lerngruppen.

EDDA kann an dieser Stelle bei der Umsetzung der im Studium erworbenen Erkenntnisse in Handlungsskripte unterstützen (im Sinne des Transformationsmodells der Unterrichtsplanung) und die Chance erhöhen, dass die Studierenden in wissenschaftlichen Kategorien und Modellen denken (im Sinne des Modells von Overmann, 1996; 2008).

Ähnlich gestaltet sich auch die Anwendung von EDDA im Vorbereitungsdienst, welcher in der strukturanalytischen Professionstheorie nach dem Studium der Einführung in die Handlungspraxis dient (vgl. Helsper, 2011, S. 150). Hier kann EDDA dazu beitragen, dass sich die Handlungspraxis nicht von den wissenschaftlichen Modellen und Konzepten entkoppelt bzw. die Verzahnung von Wissenschaft und Praxis weiter stärken. Die Ausbildungslehrkräfte und Lehrkräfte im Vorbereitungsdienst (LiV) können die Entwürfe hinsichtlich vorher festgelegter Gestaltungskriterien auswerten, die aufgrund der Implementierung von EDDA stets explizit den Begriffen wissenschaftlicher Modelle entsprechen. Dies steigert die Chance, dass auch die Besprechung und Überarbeitung der Entwürfe entlang wissenschaftlicher Kategorien erfolgt und nicht zu stark an alltagsweltlichen Konzepten oder subjektiven Theorien ausgerichtet wird. Die Wahl der Kriterien kann auch auf die individuellen Zielvereinbarungen der LiV nach den Unterrichtsbesuchen abgestimmt werden. Durch das wiederholte Hochladen kontinuierlich überarbeiteter Entwürfe kann der eigene Lernprozess bei der Unterrichtsplanung dokumentiert, reflektiert und so ein systematischer, auf wissenschaftlichen Modellen basierter Prozess zur Erreichung der Zielvereinbarungen realisiert werden. Ausbildungslehrkräfte können die ohnehin schon knappen Kapazitäten für eine effizientere Beratung nutzen, die gleichzeitig durch die Zieldimensionen der LiV effektiver für die Beförderung der professionellen Handlungskompetenz wird.

Sowohl beim Einsatz im Studium als auch im Vorbereitungsdienst sollte darauf geachtet werden, die Beziehung zwischen Lehrenden bzw. Mentor*innen und den

Studierenden bzw. LiV weiterhin zu stärken, z. B. durch Beratungstätigkeiten. Denn im Sinne der Augmentation (Jarrahi et al., 2023; Seufert et al., 2021) sollten KI-Tools zu keinem Zeitpunkt die Funktion einer Lehrkraft oder von Mentor*innen ersetzen, sondern diese, analog zum Einsatz von KI-Technologien im Schulunterricht, entlasten. Damit entspricht EDDA der Idee, sowohl Lehrende als auch Studierende im Lernprozess zu unterstützen und über Unterrichtsplanung ins Gespräch zu kommen. Abb. 14 illustriert unser Verständnis von EDDA als Ergänzung der Beziehung zwischen Lehrenden und Studierenden.

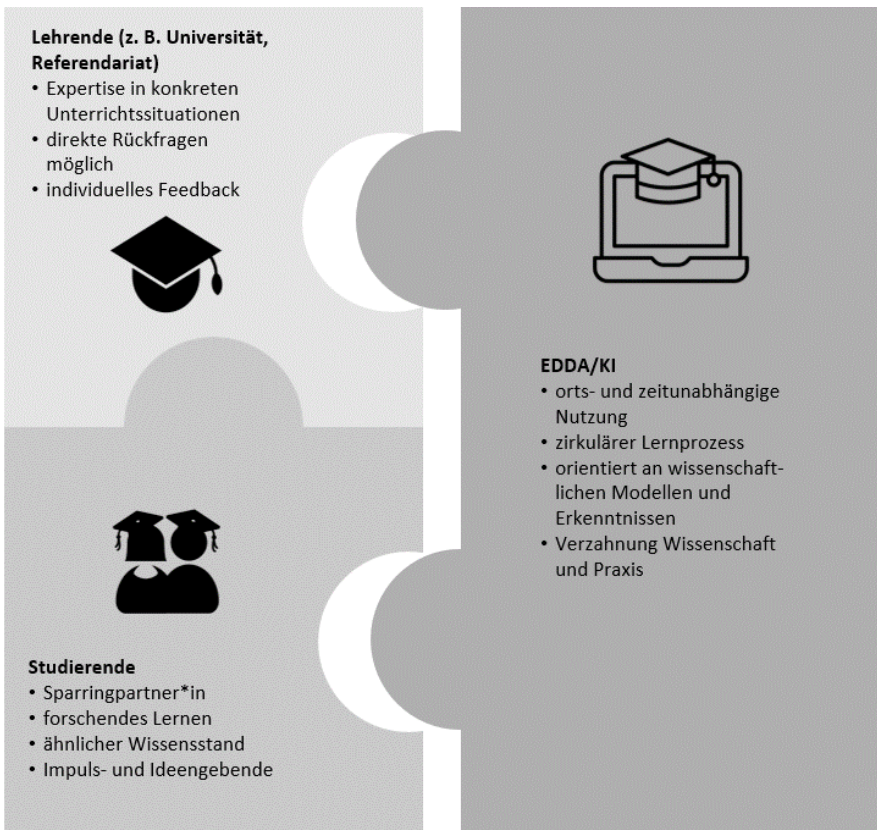


Abbildung 14: Feedbackgebende im Unterrichtsplanungsprozess in der ersten und zweiten Phase der Lehrkräftebildung

Für Lehrkräfte in der Berufstätigkeit steht mit EDDA die Möglichkeit zur kontinuierlichen Fortbildung zur Verfügung, da die KI in ihren Empfehlungen neue For-

schungsergebnisse berücksichtigt und zur Integration dieser in den Unterricht einlädt. Außerdem können Lehrkräfte neue didaktische und methodische Impulse erhalten und diese auch in ihre Unterrichtsmaterialien integrieren (Stichwort Nachhaltigkeit oder Digitalisierung).

Eine zentrale Limitation von EDDA liegt in der derzeitigen Beschränkung der Analysen auf die in Abschnitt 3.1 entwickelten Kriterien. Eine Erweiterung der Analysen und Empfehlungen, z. B. auf die curriculare und Sachanalyse sowie die Umsetzung der Lernziele befindet sich aktuell noch in der Entwicklungsphase. Kritisch ist auch anzumerken, dass EDDA nicht die gesamte Bandbreite an wissenschaftlichen Modellen zur Analyse und Rückmeldung nutzt, d. h. die Nutzenden vor allem im Umgang mit den implementierten wissenschaftlichen Modellen vertraut gemacht werden.

4.2 Ausblick: Einsatz in der Forschung

Für die Forschung bieten KI-Assistenzsysteme wie EDDA eine Infrastruktur, die es erlaubt, KI transparent aus bildungswissenschaftlicher Perspektive zu verschiedenen Fragestellungen zu beleuchten, wie:

- ▶ Gestaltung der Zusammenarbeit von Nutzenden und KI-Systemen
- ▶ Reaktionen und Nutzung von Lernenden im Umgang mit den Rückmeldungen der KI
- ▶ Veränderungen von Einstellungen und Akzeptanz gegenüber KI
- ▶ Lernzuwächse durch KI
- ▶ Einfluss von KI-Assistenzsystemen auf die Feedbackgespräche zwischen Seminarleitungen und Studierenden bzw. LiVs

Herausforderungen in der Entwicklung einer solchen Assistenz bestehen noch im hohen personellen Aufwand bei der Generierung der notwendigen Daten für das überwachte maschinelle Lernen. Diese Aufwendungen sind nur dann zu rechtfertigen, wenn eine KI wie EDDA einer großen Anzahl an Personen zugänglich gemacht und von diesen genutzt wird. So können die antrainierten Klassifizierer über das Package AI for Education (Berding et al., 2023) von Forschenden genutzt werden, um schnell große Datenmengen analysieren zu lassen, z. B. im Rahmen von Studien zur Unterrichtsplanungskompetenz, in Schulbuchanalysen oder Studien über die Wirkung von Planungen auf den realen Outcome bei Lernenden. Die Nutzung der trainierten Klassifizierer erlaubt es, personale Ressourcen bei der Analyse einzusparen, die für andere Datenerhebungen und Datenauswertungen genutzt werden können. Damit erlaubt der Tausch solcher Modelle in einer Disziplin ganz neue bzw. umfangreichere Studien.

Literatur

- Aebli, H. (2011). *Zwölf Grundformen des Lehrens* (14. Aufl.). Stuttgart: Klett-Cotta.
- Alpaydin, E. (2019). *Maschinelles Lernen* (2. Aufl.). Berlin: De Gruyter.
- Apra, C. (2018). *Instructionsdesign und Unterrichtsplanung*. In H. Niegemann & A. Weinberger (Hrsg.), *Lernen mit Bildungstechnologien* (S. 1-19). Wiesbaden: Springer.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9 (4), 469-520. <https://www.doi.org/10.1007/s11618-006-0165-2>
- Berding, F., Pargmann, J., Riebenbauer, E., Rebmann, K. & Slopinski, A. (2023). *AI for Education (aifeduction)*. A R package for educators and researchers of the educational and social sciences. Available under: <https://CRAN.R-project.org/package=aifeduction> [13.03.2024].
- Berding, F. & Pargmann, J. (2022). *Iota Reliability Concept of the Second Generation*. Logos. <https://www.doi.org/10.30819/5581>
- Berding, F., Slopinski, A. & Frerichs, R. (2020). Auszubildende als zukünftige Change Agents for Sustainable Innovations. *Betriebswirtschaftliche Forschung und Praxis*, 72 (3), 313-337.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A. & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79 (3), 1243-1289. <https://www.doi.org/10.3102/0034654309333844>
- Bundesinstitut für Berufsbildung (BIBB) (Hrsg.). (2021). *Vier sind die Zukunft*. Verfügbar unter: <https://www.bibb.de/dienst/veroeffentlichungen/de/publication/download/17281> [13.03.2024].
- Bloemen, A. (2011). *Lernaufgaben in Schulbüchern der Wirtschaftslehre*. München: Hampp.
- Brown, J. S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18 (1), 32. <https://www.doi.org/10.2307/1176008>
- Buxmann, P. & Schmidt, H. (2021). *Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens*. In P. Buxmann & H. Schmidt (Hrsg.), *Künstliche Intelligenz* (S. 3-26, 2. Aufl.). Berlin: Springer.
- Chan, B., Schweter, S. & Möller, T. (2020). German's next language model. In D. Scott, N. Bel & C. Zong (Hrsg.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 6788-6796). International Committee on Computational Linguistics. <https://www.doi.org/10.18653/v1/2020.coling-main.598>
- Chollet, F., Kalinowski, T. & Allaire, J. J. (2022). *Deep learning with R* (2nd ed.). Manning. Available under: <https://learning.oreilly.com/library/view/-/9781633439849/?ar> [13.03.2024].
- Chowdhary, K. (2020). *Fundamentals of artificial intelligence*. New Delhi: Springer. <https://www.doi.org/10.1007/978-81-322-3972-7>
- Davenport, T. & Kirby, J. (2016). Just how smart are smart machines? *MIT Sloan Management Review*, 57 (3), 21-25.
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39 (2), 223-238.
- Dörner, D. (1979). *Problemlösen als Informationsverarbeitung* (2. Aufl.). Stuttgart: Kohlhammer.
- Görz, G., Braun, T. & Schmid, U. (2021). Einleitung. In G. Görz, U. Schmid & T. Braun (Hrsg.), *Handbuch der Künstlichen Intelligenz* (S. 1-26, 6. Aufl.). De Gruyter. <https://www.doi.org/10.1515/9783110659948>
- Gwet, K. L. (2014). *Handbook of inter-rater reliability* (4th ed.). Gaithersburg: Advanced Analytics.
- Hacker, W. & Sachse, P. (2014). *Allgemeine Arbeitspsychologie*. Göttingen: Hogrefe.
- Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuan Yue, H. & Bing, G. (2017). Learning from class-imbalanced data. *Expert Systems with Applications*, 73, 220-239. <https://doi.org/10.1016/j.eswa.2016.12.035>
- Helsper, W. (2011). Lehrerprofessionalität. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 149-170). Münster: Waxmann.
- Jarrahi, M. H., Askay, D., Eshraghi, A. & Smith, P. (2023). Artificial intelligence and knowledge management. *Business Horizons*, 66 (1), 87-99. <https://www.doi.org/10.1016/j.bushor.2022.03.002>
- Jeon, J. & Lee, S. (2023). Large language models in education. *Education and Information Technologies*, 28, 1-20. <https://www.doi.org/10.1007/s10639-023-11834-1>

- Khurana, D., Koli, A., Khatteer, K. & Singh, S. (2023). Natural language processing. Multimedia tools and applications, 82 (3), 3713-3744. <https://www.doi.org/10.1007/s11042-022-13428-4>
- Kleesiek, J., Murray, J. M., Strack, C., Kaissis, G. & Braren, R. (2020). Wie funktioniert maschinelles Lernen? Der Radiologe, 60 (1), 24-31. <https://www.doi.org/10.1007/s00117-019-00616-x>
- KMK (2019). Standards für die Lehrerbildung: Bildungswissenschaften. Verfügbar unter: https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf [13.03.2024].
- Krämer, H., Jordanski, G. & Goertz, L. (2017). Medien anwenden und produzieren. Bonn: Bundesinstitut für Berufsbildung (BIBB).
- Krippendorff, K. (2019). Content Analysis (4th ed.). Thousand Oaks: Sage.
- Lane, H., Howard, C. & Hapke, H. M. (2019). Natural language processing in action. Shelter Island: Manning.
- Lanquillon, C. (2019). Grundzüge des maschinellen Lernens. In S. Schacht & C. Lanquillon (Hrsg.), Blockchain und maschinelles Lernen (S. 89-142). Berlin: Springer.
- Latif, E., Mai, G., Nyaaba, M., Wu, X., Liu, N., Lu, G., Li, S., Liu, T. & Zhai, X. (2023). Artificial General Intelligence (AGI) for Education. Preprint. Available under: <https://arxiv.org/abs/2304.12479> [13.03.2024].
- Oevermann, U. (1996). Theoretische Skizze einer revidierten Theorie professionalisierten Handelns. In A. Combe & W. Helsper (Hrsg.), Pädagogische Professionalität (S. 70-82). Frankfurt am Main: Suhrkamp.
- Oevermann, U. (2008). Profession contra Organisation. In W. Helsper, S. Busse, M. Hummrich & R.T. Kramer (Hrsg.), Pädagogische Professionalität in Organisationen (S. 55-77). Wiesbaden: VS Verlag für Sozialwissenschaften. https://www.doi.org/10.1007/978-3-531-90777-2_4
- OpenAI. (2023). GPT-4 Technical Report. Preprint. Available under: <https://arxiv.org/abs/2303.08774> [13.03.2024].
- Piaget, J. (1977). The development of thought. University of California: Viking.
- Reinisch, H. (2021). Unterrichtsplanung im Kontext bildungstheoretischer Grundannahmen. In J. Klusmeyer & M. Söll (Hrsg.), Unterrichtsplanung in der Wirtschaftsdidaktik (S. 17-34). Wiesbaden: Springer.
- Rich, E. (1983). Artificial Intelligence. New York City: McGraw-Hill.
- Riebenbauer, E. (2022). Kompetenzentwicklung im Masterstudium Wirtschaftspädagogik. Bielefeld: wbv. <https://www.doi.org/10.3278/9783763970216>
- Russel, S. J. & Norvig, P. (2021). Artificial Intelligence. (4th ed.). Harlow: Pearson.
- Salas-Pilco, S., Xiao, K. & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education. Education Sciences, 12 (8), 569. <https://www.doi.org/10.3390/educsci12080569>
- Schmaltz, C. (2019). Heterogenität als Herausforderung für die Professionalisierung von Lehrkräften. Wiesbaden: Springer. <https://www.doi.org/10.1007/978-3-658-23020-3>
- Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Wade, C. A., Surkes, M. A. & Lowerison, G. (2009). Technology's effect on achievement in higher education. Journal of Computing in Higher Education, 21 (2), 95-109. <https://www.doi.org/10.1007/s12528-009-9021-8>
- Schrader, J. & Schöb, S. (2016). Die Planung von Lehr-Lern-Einheiten mit digitalen Medien. Zeitschrift für Weiterbildungsforschung – Report, 39 (3), 331-347. <https://www.doi.org/10.1007/s40955-016-0078-5>
- Seufert, S., Guggemos, J. & Ifenthaler, D. (2021). Zukunft der Arbeit mit intelligenten Maschinen. Zeitschrift für Berufs- und Wirtschaftspädagogik (Beiheft), 31, 9-27.
- Söll, M. & Klusmeyer, J. (2018). Unterrichtsplanung als zentraler Gegenstand der Wirtschaftsdidaktik. In P. T. Tramm, M. Casper & T. Schlömer (Hrsg.), Berichte zur beruflichen Bildung (S. 73-88). Bielefeld: Bertelsmann.
- Stender, A., Brückmann, M. & Neumann, K. (2015). Vom Professionswissen zum kompetenten Handeln im Unterricht: Die Rolle der Unterrichtsplanung. Beiträge zur Lehrerinnen- und Lehrerbildung, 33 (1), 121-133.
- Stütz, S., Berding, F., Reincke, S. & Scheper, L. (2022). Characteristics of learning tasks in accounting textbooks. Empirical Research in Vocational Education and Training, 14 (1), 1-33. <https://www.doi.org/10.1186/s40461-022-00138-2>

- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C. & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning. *Review of Educational Research*, 81 (1), 4-28. <https://www.doi.org/10.3102/0034654310393361>
- Tunstall, L., von Werra, L., Wolf, T. & Géron, A. (2022). *Natural language processing with transformers* (Rev. ed.). Newton: O'Reilly.
- United Nations (2015). *Transforming our world: The 2030 agenda for sustainable development*. Available under: <https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf> [13.03.2024].
- van der Kleij, F. M., Feskens, R. C. W. & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes. *Review of Educational Research*, 85 (4), 475-511. <https://www.doi.org/10.3102/0034654314564881>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. (2017). Attention is all you need. Preprint. Available under: <https://arxiv.org/abs/1706.03762v7> [13.03.2024].
- Welsch, A., Eitle, V. & Buxmann, P. (2018). Maschinelles Lernen. *HMD Praxis der Wirtschaftsinformatik*, 55 (2), 366-382. <https://www.doi.org/10.1365/s40702-018-0404-z>
- Wesche, J. S. & Sonderegger, A. (2019). When computers take the lead. *Computers in Human Behavior*, 101, 197-209. <https://www.doi.org/10.1016/j.chb.2019.07.027>
- Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M. & Drachsler, H. (2021). Are we there yet? – A systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence*, 4, 654924. <https://www.doi.org/10.3389/frai.2021.654924>
- Wong, J., Baars, M., Koning, B. B. de, van der Zee, T., Davis, D., Khalil, M., Houben, G.-J. & Paas, F. (2019). Educational theories and learning analytics: From data to knowledge. In D. Ifenthaler, D.-K. Mah & J. Y.-K. Yau (Hrsg.), *Utilizing learning analytics to support study success* (S. 3-25). Cham: Springer. https://www.doi.org/10.1007/978-3-319-64792-0_1

Autor*innen

Julia Pargmann, wiss. Mitarbeiterin, Berufliche Bildung und Lebenslanges Lernen, Universität Hamburg

Anna Leube, wiss. Mitarbeiterin, Berufliche Bildung und Lebenslanges Lernen, Universität Hamburg

Prof. Dr. Florian Berding, Berufliche Bildung und Lebenslanges Lernen, Universität Hamburg

Assoz. Prof. Dr. Elisabeth Riebenbauer, Institut für Wirtschaftspädagogik, Universität Graz

Prof. Dr. Karin Rebmann, Fachgebiet Berufs- und Wirtschaftspädagogik, Universität Oldenburg

Dr. Andreas Slopinski, Fachgebiet Berufs- und Wirtschaftspädagogik, Universität Oldenburg

Michael Gillert, wiss. Mitarbeiter, Berufliche Bildung und Lebenslanges Lernen, Universität Hamburg

Korrespondenz an: julia.pargmann@uni-hamburg.de

Allgemeiner Teil

Joel Guttke & Raphaela Porsch

Kognitive Interviews als Methode der Inhaltsvalidierung von Fragebogenitems zur Erfassung kognitiver Aktivierung im Englischunterricht der Grundschule

Schüler- und Lehrkräftebeurteilungen gelten als belastbare Datenquelle in der Unterrichtsforschung. Lenseke (2016) weist jedoch nach, dass Grundschüler*innen Fragebogenitems zu Unterrichtsqualität – u. a. zu Aktivierung – häufig nicht im intendierten Sinne bearbeiten. Dieser Befund verdeutlicht die Notwendigkeit eines expliziten Nachweises der Inhaltsvalidität in der Fragebogenkonstruktion. Für die Inhaltsvalidität von Lehrkräfteurteilen zu Unterrichtsqualität liegen bislang keine Befunde vor. Der Beitrag präsentiert Ergebnisse einer qualitativen Studie zur Sicherung der Inhaltsvalidität neukonstruierter Items zur fachspezifischen Erfassung kognitiver Aktivierung im Englischunterricht der Grundschule. In kognitiven Interviews wurden neun Englischlehrerinnen und 39 Grundschüler*innen der Klasse 4 zu Fragebogenitems befragt und die Interviewtranskripte mittels strukturierender qualitativer Inhaltsanalyse ausgewertet. Damit repliziert die Studie den Ansatz von Lenseke (2016) für das Fach Englisch und ergänzt ihn um die Gruppe der Lehrkräfte. Die Ergebnisse belegen, dass die Mehrheit der Items in beiden Probandengruppen nicht im intendierten Sinn bearbeitet wurde. Davon ausgehend werden Strategien zur Sicherung der Inhaltsvalidität formuliert.

Schlagwörter: Fragebogenentwicklung – kognitive Aktivierung – kognitives Interview – Unterrichtsqualität

Assessing the content validity of questionnaire items on cognitive activation in English language teaching at primary level through cognitive interviewing

Student and teacher feedback on instructional quality is considered a tried and tested source of data. However, Lenseke (2016) demonstrates that students at primary level often fail at understanding and interpreting questionnaire items on instructional quality – student activation, in particular – as intended. These findings emphasize the need for explicitly addressing issues of content validity in student questionnaires. Studies on the content validity of teacher ratings of instructional quality have not been conducted yet. This paper presents a qualitative study investigating the content validity of newly constructed items on cognitive activation in English language teaching at primary level. Cognitive interviews were conducted with nine primary English language teachers and 39 fourth year students to explore how they process questionnaire items on cognitive activation. The interview data was analyzed using qualitative content analysis. Thus, the study replicates Lenseke's (2016) approach and adapts it to teachers as a new target group. The findings show that both groups of participants misinterpreted more than half of the items each. Building on these findings, the paper suggests strategies for ensuring the content validity of items.

Keywords: cognitive activation – cognitive interviewing – instructional quality – questionnaire design

1 Einleitung

Der Annahme von Angebots-Nutzungs-Modellen schulischer Lehr-Lern-Prozesse folgend nimmt Unterrichtsqualität eine vermittelnde Rolle zwischen Lehrkräftemerkmalen und Lernergebnissen von Schüler*innen ein (Vieluf, Praetorius, Rakoczy, Kleinknecht & Pietsch, 2020). Unterrichtsqualität beschreibt jene Unterrichtsmerkmale, „die nachweislich mit einer Entwicklung der Lernenden im Sinne der Realisierung von Bildungs- und Erziehungszielen einhergehen“ (Klieme, 2022, S. 414). Zur Systematisierung von Unterrichtsqualitätsmerkmalen existiert im deutschsprachigen Raum das Modell der drei Basisdimensionen und ein – vergleichsweise junges und folglich wenig empirisch beforschtes – Syntheseframework (Praetorius, Rogh & Kleickmann, 2020). Das Modell der drei Basisdimensionen von Unterrichtsqualität unterscheidet drei lernwirksame unterrichtliche Angebotsstrukturen: Klassenführung, kognitive Aktivierung und konstruktive Unterstützung. Klassenführung und konstruktive Unterstützung beschreiben Maßnahmen zur Erzielung eines hohen Maßes an time on task und Motivation sowie eines positiven Selbstkonzepts. Kognitiv aktivierender Unterricht zeichnet sich durch Merkmale aus, denen das Potenzial zugeschrieben wird, eine elaborierte Verarbeitung der Unterrichtsinhalte seitens der Schüler*innen zu fördern (Lipowsky, 2020). Dazu zählt beispielsweise die Aufgabenkomplexität oder der Einbezug von Schülervorstellungen (Klieme, 2022).

Um die drei Basisdimensionen empirisch zugänglich zu machen, werden Unterrichtsbeurteilungen durch externe Rater, Lehrkräfte und Schüler*innen mittels Fragebögen als Datenquelle herangezogen. Dass zwischen den drei Perspektiven auf Unterricht wenig Einigkeit herrscht, wurde empirisch anhand geringer Korrelationen wiederholt nachgewiesen (Clausen, 2002; Fauth, Decristan, Rieser, Klieme & Büttner, 2014a, b). Begründet wird die Perspektivendivergenz entlang zweier Argumentationslinien: Einerseits unterliegen alle drei Perspektiven Urteilsfehlern, welche die Urteilsvalidität einschränken (Lenske, 2016; Lenske & Praetorius, 2020). Es wird folglich bei jeder Akteursgruppe von einem perspektivenspezifischen Messfehler ausgegangen. Andererseits zeigen Kunter und Baumert (2006), dass sich die Akteursgruppen aufgrund der perspektivenspezifischen Validität ihrer Urteile unterschiedlich gut dazu eignen, verschiedene Unterrichtsqualitätsmerkmale zu beurteilen. Anhand von PISA-Daten aus dem Jahr 2003 arbeiten die Autor*innen heraus, dass Lehrkräfte- und Schülerurteilen zu Unterrichtsqualität unterschiedliche Faktorenstrukturen unterliegen und die Urteile beider Akteursgruppen nur hinsichtlich Klassenführung und Interaktionstempo übereinstimmen (Kunter & Baumert, 2006).

Die Bedeutung von Items zur Erfassung von Unterrichtsqualität wird in bisherigen Untersuchungen zu Perspektivunterschieden wenig diskutiert. Solche Items enthalten Informationen über Raterperspektive (Wer bearbeitet das Item?) und Itemreferenz (Auf wen bezieht sich das Item?; Fauth, Göllner, Lenske, Praetorius & Wagner,

2020). Daraus lässt sich schließen, wie hoch der Grad der Selbstbeurteilung ausfällt und inwiefern den Adressat*innen ausreichend Informationen vorliegen, um das Item zu bearbeiten. Im Falle kognitiver Aktivierung unterscheiden Fragebogenitems bspw. häufig nicht zwischen dem angebotsseitigen Potenzial zu kognitiver Aktivierung und der kognitiven Aktivität der Schüler*innen (Rieser & Decristan, 2023).

Neu formulierte Items sind daher im Prozess der Fragebogenentwicklung zwingend auf ihre Verständlichkeit und damit auf Inhaltsvalidität zu prüfen. Der Beitrag verfolgt das Ziel, die Inhaltsvalidität neukonstruierter Schüler- und Lehrkräfteitems zur Erfassung kognitiver Aktivierung im Englischunterricht der Jahrgangsstufe 4 anhand kognitiver Interviews zu prüfen. Damit repliziert er das Vorgehen von Lenske (2016) konkret für den Englischunterricht und erweitert es um die Gruppe der Lehrkräfte.

2 Forschungsstand

2.1 Kognitive Aktivierung im Fach Englisch

Im Diskurs über Unterrichtsqualität ist seit geraumer Zeit eine Aushandlung über die Fachspezifik kognitiver Aktivierung zu beobachten. Es wird diskutiert, inwiefern bestehende Modelle der Unterrichtsqualität fachspezifisch ausdifferenzieren (Praetorius, Herrmann, Gerlach, Zülsdorf-Kersting, Heinitz & Nehring, 2020b) oder durch weitere fachspezifische Dimensionen zu ergänzen seien (Lipowsky, Drollinger-Vetter, Klieme, Pauli & Reusser, 2018). So besteht Zweifel, „ob die kognitive Aktivierung fach- und inhaltspezifische Facetten der Unterrichtsqualität differenziert genug einzufangen vermag“ (Lipowsky & Bleck, 2019, S. 228). Dieser Beitrag basiert auf der Annahme, dass fachspezifische Indikatoren dazu beitragen können, kognitive Aktivierung im Englischunterricht valider zu erfassen.

Obwohl kognitive Aktivierung in der Fremdsprachendidaktik im deutschsprachigen Raum bisher nur selten beforscht wurde (Guttke, 2023), lassen sich in der Literatur zu Instructed Second Language Acquisition (Loewen, 2020) fachspezifische Aspekte des Konstrukts aus vier Bereichen identifizieren: (1) Target language input, (2) target language output, (3) corrective feedback und (4) information processing. Die in der durchgeführten Studie entwickelten Items bilden diese vier Bereiche ab, weshalb sie knapp erläutert werden.

Hinsichtlich des lehrkräfteseitigen Angebots von target language input (Bereich 1) gilt es zu berücksichtigen, dass die englische Sprache nicht nur Gegenstand, sondern auch Medium von Lehr-Lernprozessen im Englischunterricht ist (Wilden, 2021). Schüler*innen setzen sich nicht ausschließlich explizit mit der Zielsprache und anglophonen Kulturen auseinander (z. B. bei der Vermittlung eines grammatischen Phänomens), sondern die Zielsprache wird vielmehr in Kommunikation mit ihrer Lehrkraft und ihren Mitschüler*innen gebraucht. Damit eine möglichst tiefe Aus-

einandersetzung mit dem Lerngegenstand überhaupt stattfindet, muss die Lehrkraft im Sinne eines Potenzials zu kognitiver Aktivierung sicherstellen, dass der target language input von den Schüler*innen verstanden wird (Kersten, 2021). Möglich ist dies bspw. durch die Nutzung multipler Kommunikationskanäle und -modi (z. B. Gestik). Analog ist es hinsichtlich der Befähigung der Schüler*innen zur Produktion von target language output (Bereich 2) unerlässlich, sie auf kognitiver (z. B. language scaffolding) und emotionaler Ebene (z. B. Ermutigung) zu unterstützen. Diese beiden Bereiche sind im Englischunterricht der Grundschule umso relevanter, da die Kommunikationsfähigkeiten der Schüler*innen in der Zielsprache dort gewöhnlich noch gering sind (Kolb & Schocker, 2021).

Dem positiven Umgang mit Fehlern – die den Zweitspracherwerbsprozess zwangsläufig begleiten – in Form von corrective feedback (Bereich 3) kommt aufgrund der soeben beschriebenen Gegebenheiten eine besondere Bedeutung zu. Lehrkräfte sind mit der Herausforderung konfrontiert, ein ausgewogenes Verhältnis zwischen focus on form und focus on fluency zu schaffen, um einerseits der Fossilisierung von Fehlern vorzubeugen und andererseits Schüler*innen nicht im Gebrauch der Zielsprache zu hemmen. Dazu gehört auch die Wahl angemessener Feedbackstrategien. Im Sinne der curricular vorgeschriebenen kommunikativen Ausrichtung des Englischunterrichts sollten Schüler*innen ausreichend Gelegenheit erhalten, die Zielsprache in bedeutsamen Kontexten zur Auseinandersetzung mit kommunikativen Problemstellungen einzusetzen (Wilden, 2021). Solche Situationen veranlassen Schüler*innen beispielsweise dazu, die situative Angemessenheit ihres zielsprachlichen outputs zu reflektieren oder Kompensationsstrategien einzusetzen, um Kommunikation aufrechtzuhalten. Damit tragen sie zu einer besonders tiefen Verarbeitung der Zielsprache (Bereich 4) bei.

2.2 Lehrkräfte- und Schülerurteile zu Unterrichtsqualität

Den drei beschriebenen Perspektiven von externen Ratern, Lehrkräften und Schüler*innen auf Unterrichtsqualität werden jeweils eigene Vor- und Nachteile zugeschrieben. Da der Beitrag die Entwicklung von Items für Schüler- und Lehrkräftefragebögen thematisiert, werden die perspektivenspezifischen Charakteristika von Schüler- und Lehrkräfteurteilen zu Unterrichtsqualität zusammengefasst, bevor der Forschungsstand zu deren (Inhalts-)Validität dargestellt wird.

Die Erfassung von Schülerurteilen zu Unterrichtsqualität mittels Fragebögen hat sich als vergleichsweise ökonomischer Zugang in der Unterrichtsqualitätsforschung etabliert. Grund dafür sind zahlreiche Vorteile, die Schülerurteilen im Vergleich zu Lehrkräfte- oder Beobachterurteilen zugeschrieben werden (Lenske & Praetorius, 2020). Schüler*innen verfügen sowohl über Langzeiterfahrung mit derselben Lehrkraft als auch über Vergleichsmöglichkeiten durch Unterricht bei verschiedenen

Lehrkräften. Dies sei der Validität sowie der zeitlichen Stabilität von Schülerurteilen zuträglich (Clausen, 2002). Zudem weisen Schülerurteile durch die Zusammenfassung von Einzelurteilen je Klasse in der Regel eine hohe Reliabilität auf (Göllner, Wagner, Klieme, Lüdtke, Nagengast & Trautwein, 2016). Schließlich seien Schülerurteile für die Erfassung kognitiver Aktivität, die eine individuelle Nutzung unterrichtlicher Lernangebote beschreibt, besonders gewinnbringend. Sie erlaubten eine direkte Auskunft über das individuelle Unterrichtserleben, das von Lehrkräften und Beobachter*innen hingegen hoch-inferent beurteilt werde (Lenske, 2016).

Lehrkräfte gelten aufgrund ihrer umfangreichen Ausbildung und Unterrichtserfahrung als Expert*innen für Lehr-Lernprozesse. Ihren Urteilen zu Unterrichtsqualität wird eine hohe Validität zugeschrieben, da das Professionswissen von Lehrkräften einen Zugang zum didaktischen Verständnis und zur Bewertung von Unterrichtsmerkmalen ermöglicht, der Schüler*innen durch einen Mangel an Expertise fehlt (De Jong & Westerhof, 2001). Zudem haben Lehrkräfte einen detaillierten Zugriff auf Informationen zum Unterrichtsgeschehen, weil sie ihren Unterricht eigens planen und durchführen (Kunter & Baumert, 2006). Schließlich ist davon auszugehen, dass Lehrkräfte das Potenzial zu kognitiver Aktivierung der von ihnen eingesetzten Lernaufgaben und des Unterrichtsdiskurses didaktisch begründen können. Ein valider Rückschluss über die individuelle Nutzung dieses Potenzials ist Lehrkräften hingegen einzig aufgrund der erworbenen Leistungen – hier: der sprachlichen Kompetenzen – ihrer Schüler*innen möglich. Darüber hinaus lässt sich das Nutzungsverhalten von Schüler*innen über distale Indikatoren, wie bspw. ihre Unterrichtsbeteiligung einschätzen (Jansen, Decristan & Fauth, 2022). Trotz ihrer Expertise unterliegen auch Lehrkräfte Urteilsfehlern, welche die Validität ihrer Urteile zu Unterrichtsqualität einschränken können (Wisniewski, Röhl & Fauth, 2022).

Grundsätzlich zeigt die Literatursichtung, dass Untersuchungen der Unterrichtsqualität mit dem Schwerpunkt auf der Prüfung der Inhaltsvalidität in begrenzter Anzahl vorliegen. Die Arbeit von Lenske (2016), in der Grundschüler*innen mithilfe kognitiver Interviews zu Itemverständnis und -interpretation untersucht wurden, zeigt eine Reihe von Schwierigkeiten der Befragten auf. Dazu zählte insbesondere, dass die Schüler*innen Begriffe in den Items nicht oder nicht in der intendierten Weise verstehen. Verständnisprobleme traten bei etwa der Hälfte der 59 getesteten Items auf. Folglich ist eine Form der Inhaltsvalidierung unter Einbezug von Schüler*innen insbesondere für neu entwickelte Items wünschenswert. Lenske und Praetorius (2020) untersuchten, inwieweit sich die eingeschränkte Validität auch bei der Untersuchung von Sekundarschüler*innen aufzeigen lässt. Die Schüler*innen nahmen an kognitiven Interviews teil und wurden zu ihrer Urteilssicherheit befragt. Die Ergebnisse zeigen, dass die Schüler*innen Items zur Erfassung der kognitiven Aktivierung sowie Differenzierung häufig nicht im intendierten Sinne verstanden, was die Notwendig-

keit einer Untersuchung der Inhaltsvalidität von Items zu diesem Konstrukt unterstreicht. Neben den Charakteristika der Items (z. B. Komplexität) und Merkmalen der Schüler*innen (z. B. Vertrautheit mit Begriffen) liegen weitere Einflüsse auf Ratings von Items zu Unterrichtsqualität vor. Beispielsweise verweisen Röhl und Rollett (2020) darauf, dass die Bewertung der Lehrkraft als freundliche und emotional warmherzige Person die Unterrichtswahrnehmung positiv beeinflusst. Auch Fauth und Kolleg*innen (2014b) weisen auf nicht zu vernachlässigende Zusammenhänge zwischen Lehrkräftepopularität und den drei Basisdimensionen hin. Deutlich wird, dass das Verständnis und Rating durch Schüler*innen einer Reihe von Einflussfaktoren unterliegen.

Studien, welche die Verständlichkeit von Lehrkräfteteams zu Unterrichtsqualität adressieren, konnten nicht identifiziert werden. Es scheint vorausgesetzt zu werden, dass entweder die Prüfung von Konstruktvalidität ausreichend ist (z. B. Wafudu, Kamin & Marcel, 2022) oder die Unterrichtserfahrung und das Professionswissen sowie die im Vergleich zu Grundschul*innen voll entwickelte Sprachkompetenz von Lehrkräften eine hinreichend inhaltssvalide Itembearbeitung sicherstellen. Diese Annahmen greifen vor dem Hintergrund empirischer Forschungsergebnisse zur Lehrkräfteprofessionalisierung jedoch zu kurz. Einerseits könnte die valide Bearbeitung von Items zu Unterrichtsqualität mit der Berufserfahrung von Lehrkräften zusammenhängen. Stufenmodellen der Lehrkräfteprofessionalisierung zufolge nehmen Noviz*innen ihren Unterricht weniger holistisch und präzise wahr, weil ihnen Erfahrungswissen und Unterrichtsroutinen fehlen (Bromme, 2014). Andererseits variiert das Professionswissen von Englischlehrkräften an der Grundschule im Vergleich zu anderen Fächern stark, da ein fundamentaler Anteil des Englischunterrichts von fachfremden Lehrkräften unterrichtet wird (Porsch & Wilden, 2022). Diese Überlegungen deuten an, dass auch bei Lehrkräften nicht grundsätzlich von einer inhaltssvaliden Bearbeitung von Items zu Unterrichtsqualität ausgegangen werden darf.

3 Forschungsfragen

In Abschnitt 2 wurde dargelegt, dass Schüler*innen einen einzigartigen Zugang zu Unterrichtsqualität bieten, die Validität ihrer Urteile jedoch stark von der Itemqualität abhängt. Im Englischunterricht der Grundschule müssen die Schüler*innen für die Itembearbeitung trotz ihrer begrenzten Fähigkeiten in der Zielsprache einen Transfer zwischen dem englischsprachig durchgeführten Unterricht und den deutschsprachig formulierten Items leisten. Obwohl für die Inhaltsvalidität von Lehrkräfteurteilen zu Unterrichtsqualität keine empirischen Befunde vorliegen, wurden begründete Annahmen über Faktoren, welche die Iteminterpretation einschränken könnten, formuliert. Um die Iteminterpretation beider Zielgruppen besser nachzuvollziehen, untersucht die vorliegende Studie folgende Fragen:

- (1) In welchem Ausmaß entspricht die mittels Interviewtranskription dokumentierte Iteminterpretation von Grundschüler*innen und Englischlehrkräften der Klasse 4 den Inhalten, die im Rahmen der Itemkonstruktion intendiert wurden?
- (2a) Welche häufigen Fehlertypen hinsichtlich der Iteminterpretation lassen sich in den Interviewdaten identifizieren?
- (2b) Welche Strategien zur Itemüberarbeitung ergeben sich aus diesen Fehlertypen?

4 Methode

Der Beitrag ist Teil eines Forschungsprojekts, das sich mit der Entwicklung und Validierung fachspezifischer Schüler- und Lehrkräftefragebögen zur Erfassung kognitiver Aktivierung im Englischunterricht der Jahrgangsstufe 4 beschäftigt. Zunächst wurde theoriebasiert auf Grundlage von Forschungsbefunden der Instructed Second Language Acquisition Research (Loewen, 2020) ein Pool analog formulierter Schüler- und Lehrkräfteitems konstruiert und mithilfe einer Online-Expert*innenbefragung ($n = 24$) hinsichtlich seiner Konstruktrelevanz und -repräsentanz in Bezug auf das Vorliegen von Inhaltsvalidität geprüft (Guttke & Porsch, im Druck). Nach der Revision lagen für den Schülerfragebogen 49 Items und für den Lehrkräftefragebogen 53 Items vor, die sich auf die vier in Abschnitt 2.1 beschriebenen Subskalen kognitiver Aktivierung verteilen (s. Tab. 1). Die Itemanzahl beider Fragebögen ist unterschiedlich, da einzelne Items konstruiert wurden, die nur aus einer Perspektive reliabel bearbeitbar sind.

4.1 Kognitive Interviews

Mit beiden Probandengruppen wurden kognitive Interviews (Collins, 2015), eine Interviewform zur Erprobung von Fragebögen, durchgeführt. Basierend auf der Annahme, dass Fragebogenitems aufgrund individueller Kontextfaktoren unterschiedlich interpretiert werden, ermöglichen kognitive Interviews, die Itemverarbeitung durch Proband*innen zu rekonstruieren (Willis, 2005). Tourangeau, Rips und Rasinski (2012) zufolge umfasst die Itemverarbeitung vier Phasen: das Verständnis von Itembedeutung und -intention, den Abruf der zur Itembeantwortung relevanten Informationen aus dem Langzeitgedächtnis, die Bewertung und Interpretation dieser Informationen sowie die Zuordnung des eigenen Urteils zu einer der dargebotenen Antwortoptionen. Kognitive Interviews haben zum Ziel, Selbstberichte der Proband*innen zu diesen vier Phasen zu elizitieren.

Tabelle 1: Beispielitems für die vier Subskalen des neu konstruierten Fragebogens zur Erfassung kognitiver Aktivierung im Englischunterricht der Primarstufe

Subskala [Item]	Lehrkräftefragebogen	Schülerfragebogen
target language input [koerpersprache]	Wenn ich mit den Schüler*innen Englisch spreche, verdeutliche ich Wörter, die zentral für das Verständnis sind, mit Körpersprache (Gestik, Mimik).	Wenn meine Englischlehrerin auf Englisch mit mir spricht, zeigt sie manche Wörter mit ihrem Gesicht oder mit Bewegungen.
information processing [lernziel]	Ich nenne den Schüler*innen zu Unterrichtsbeginn das Lernziel der Unterrichtsstunde.	Meine Englischlehrerin sagt mir am Anfang der Unterrichtsstunde, was ich lernen werde.
target language output [wartezeit]	Nach einer Frage oder Aufforderung bin ich geduldig und gebe den Schüler*innen Zeit zu reagieren.	Wenn meine Englischlehrerin mir eine Frage stellt, gibt sie mir genügend Zeit, etwas zu sagen.
corrective feedback [explicit_correction]	Wenn Schüler*innen einen Fehler machen, benenne ich den Fehler explizit und korrigiere ihn daraufhin.	Wenn ich auf Englisch etwas falsch sage, sagt mir meine Englischlehrerin, dass ich einen Fehler gemacht habe und korrigiert den Fehler.

Das laute Denken und verbal probing zählen zu den zwei etabliertesten Techniken kognitiver Interviews (Willis, 2005). Beim lauten Denken werden die Proband*innen dazu aufgefordert, ihre Gedanken während der Itembearbeitung zu artikulieren. Einerseits ermöglicht lautes Denken den Proband*innen, ihnen relevante Itemaspekte zu fokussieren. Andererseits birgt die persönliche Relevanzsetzung die Gefahr, Aspekte zur Beantwortung der Forschungsfrage zu verfehlen. Aus diesem Grund wird lautes Denken häufig mit verbal probing kombiniert, wozu den Proband*innen spezifische Nachfragen zu Items oder ihren Antworten darauf gestellt werden. Verbal probes werden gezielt eingesetzt, um für das Forschungsinteresse relevante Aspekte zu fokussieren oder Aussagen aus dem lauten Denken zu prüfen (Collins, 2015). Die beiden Interviewtechniken ergänzen sich komplementär; sie sollten stets so implementiert werden, dass sie dem Erkenntnisinteresse dienlich sind.

4.2 Stichprobe & Datenerhebung

Die Daten stammen aus einer Gelegenheitsstichprobe von 39 Grundschüler*innen der Jahrgangsstufe 4 aus drei Grundschulen und neun Englischlehrerinnen aus neun Grundschulen in Nordrhein-Westfalen. Bei der Rekrutierung wurde darauf geachtet, dass nicht alle Proband*innen einer Klasse/Schule angehören, um eine Variation an Unterrichtssituationen sicherzustellen. Die Englischlehrerinnen weisen ein durchschnittliches Alter von 35 Jahren und eine durchschnittliche Berufserfahrung von acht Jahren (ohne Vorbereitungsdienst) vor. Sieben der Englischlehrerinnen haben das Fach Englisch auf Lehramt für die Primarstufe studiert; die beiden anderen Lehrerinnen haben die Lehrbefähigung für das Fach Englisch im Rahmen einer Nachqualifizierung erhalten. Unter den 39 Grundschüler*innen befanden sich 22 weibliche Probandinnen. Die Grundschulen der Schülerstichprobe gehörten den Sozialindexstufen eins ($n = 19$), drei ($n = 9$) und sieben ($n = 11$) an.

Alle Interviews wurden an den Schulen oder vereinzelt über das Videokonferenztool Zoom durchgeführt und audiografiert. Im Falle der Lehrerinnen handelte es sich um ein halbstandardisiertes Leitfadeninterview (durchschnittliche Dauer: 57 Min.) mit Anteilen un gelenkten lauten Denkens sowie verbal probing. Das laute Denken wurde den Lehrerinnen zu Beginn des Interviews demonstriert und anhand einer Beispielfrage geübt (s. Anhang A). Im Falle der Schüler*innen handelte es sich um ein vollstandardisiertes Leitfadeninterview (durchschnittliche Dauer: 9 Min.), im Rahmen dessen unterschiedliche verbal probing Techniken eingesetzt wurden, um verbale Berichte über den Interpretationsvorgang der Items zu elizitieren (s. Anhänge B, C). Dabei wurde auf den Einsatz un gelenkten lauten Denkens verzichtet, da es vor dem Hintergrund des Entwicklungsstands der Schüler*innen als zu komplex eingeschätzt wurde (Emde & Fuchs, 2012). Die Erziehungsberechtigten der Schüler*innen wurden um eine schriftliche Einwilligungserklärung gebeten. Die Interviews mit den Schüler*innen fanden jeweils im Anschluss an eine durch den Erstautor beobachtete Englischstunde statt.

Während der Interviews wurde eine zufällige Auswahl von durchschnittlich fünf Items (Schülerinterviews) bzw. 18 Items (Lehrkräfteinterviews) schriftlich präsentiert und vorgelesen, da eine Bearbeitung des gesamten Itempools nicht zumutbar gewesen wäre. Jedes Item wurde von mindestens drei Proband*innen bearbeitet. Nach der Präsentation eines Items wurden die Proband*innen zur Auswahl und Begründung einer Antwort (Stimme nicht zu – Stimme weniger zu – Stimme eher zu – Stimme voll zu) aufgefordert. In den Schülerinterviews wurde die Begründung unmittelbar durch cognitive probes eliziert; in den Lehrkräfteinterviews erfolgten die cognitive probes erst im Anschluss an das laute Denken.

Die Formulierung und der Einsatz von cognitive probes erfolgte in beiden Probandengruppen systematisch. Dazu wurde das Question Pitfalls Modell (Willis, 2005, S. 79-84) herangezogen. Das Modell differenziert zwischen sieben möglichen Fehlerquellen und dient dazu, Items nach der Konstruktion auf potenzielle Herausforderungen bezüglich der in Abschnitt 4.1 geschilderten Phasen der Itemverarbeitung zu untersuchen: (1) Vorlesen der Frage durch den Interviewer, (2) Instruktionen während des Interviews, (3) Verständlichkeit der Items, (4) in Items (implizit) enthaltene Präsuppositionen, (5) Verfügbarkeit oder Abruf von Wissen zur Beantwortung von Items, (6) Vertraulichkeit der durch ein Item elizitierten Informationen und (7) Passung zwischen Iteminhalt und Antwortkategorien. Zur Formulierung der cognitive probes wurde für jedes Item geprüft, welche der sieben Fehlerquellen potenziell auf das Item zutreffen könnten. Davon ausgehend wurden Testziele und cognitive probes zur Annäherung an diese Testziele formuliert (s. Anhang A).

Zur Illustration des Vorgehens dient das Lehrkräftemitem [rituale]: „Es gibt regelmäßige Sprechaktivitäten zur Wiederholung zentraler sprachlicher Elemente“. Für dieses Item wurden mit Hilfe des Question Pitfalls Model zwei Herausforderungen hinsichtlich der Itemverständlichkeit und des Informationsabrufs identifiziert. Erstens war es von Interesse, wie die Englischlehrerinnen „Sprechaktivitäten“ und „zentrale sprachliche Elemente“ definieren. Dazu wurde eine probe zur Paraphrasierung eingesetzt: „Bitte wiederholen Sie das Item in Ihren eigenen Worten“. Zweitens sollte der Informationsabruf zur Regelmäßigkeit der Sprechaktivitäten geprüft werden. Dazu wurde folgende probe eingesetzt: „Wann haben Sie [Sprechaktivität aus Antwort der Lehrkraft] das letzte Mal im Englischunterricht der Jahrgangsstufe 4 eingesetzt?“.

4.3 Datenauswertung

Die Audiomitschnitte der Interviews wurden durch Transkription in Schriftdeutsch übertragen. Auf eine Transkription der Paralinguistik wurde verzichtet, da diese Informationen vor dem Hintergrund der Forschungsfragen nicht als bedeutungsstragend erachtet wurden.

Um die schriftlich vorliegenden Daten in nominale Daten zu transformieren, wurde mittels der frei verfügbaren Software QCMap (<https://www.qcmap.org>) eine skalierend-strukturierende qualitative Inhaltsanalyse (Mayring, 2022) durchgeführt. Die einzelnen Interviews bildeten dazu jeweils die Auswertungseinheit. Innerhalb dieser wurden Probandenäußerungen in Form eines vollständigen Satzes als Kodiereinheit und die Gesamtheit aller Sprecher*innenwechsel zu einem Item als Kontexteinheit definiert. Auf Grundlage des in Abschnitt 2.2 geschilderten Forschungsstandes wurde ein deduktives Kategoriensystem mit vier Oberkategorien formuliert (s. Anhang D). In den meisten Fällen fiel die Kodierung der Probandenäußerungen

einheitlich aus, sodass die Einzelkodierungen dem Globalurteil entsprachen. In Ausnahmen konkurrierten Kodierungen unterschiedlicher Oberkategorien miteinander. Diese Fälle wurden separat erneut konsensuell kodiert.

Die Kodierung des Materials erfolgte durch den Erstautor und eine geschulte Raterin. Die Interrater-Übereinstimmung (s. Tab. 2) fällt akzeptabel aus (Caspar & Wirtz, 2002; Krippendorff, 2019).

Tabelle 2: Maße der Interrater-Übereinstimmung für Kodierung der Passung zwischen dokumentierter und intendierter Iteminterpretation in den Lehrkräfte- und Schülerinterviews

Maß	Lehrkräftedaten (n = 9)	Schülerdaten (n = 39)
Konsensmaße		
Prozentuale Übereinstimmung	.75	.79
Cohens Kappa	.58	.69
Krippendorffs Alpha	.80	.85
Konsistenzmaß		
Kendalls Tau	.75	.81

Anmerkungen: n = Stichprobengröße

5 Ergebnisse

5.1 Rating der Passung zwischen intendierter und dokumentierter Iteminterpretation

Im Falle des Lehrkräftefragebogens (s. Abb. 2, Anhang E) zeigte sich bei 18 aus 53 Items eine vollständige Passung zwischen intendierter und dokumentierter Iteminterpretation. Diese Items wurden von allen Probandinnen im intendierten Sinn beantwortet. Darunter befinden sich sowohl Items, welche die Gestaltung des language input der Englischlehrerinnen in der Zielsprache thematisieren, als auch solche, die Anlass zur Reflexion über Aufgabenauswahl und -implementation im Englischunterricht geben. Zur Illustration sei ein Ausschnitt aus dem Interview mit Lehrerin 2 zu dem Item [aufgabenbedeutung] („Ich stelle sicher, dass die Schüler*innen wissen, wie ihnen die zu bearbeitenden Aufgaben dabei helfen, das Lernziel der Unterrichtsstunde zu erreichen.“) angeführt:

L2: Stimme eher zu. Also generell ist mir schon wichtig, dass die Kinder wissen, warum machen wir diese Aufgabe denn jetzt? Was ist unser Ziel dabei? Aber, wenn man ganz ehrlich ist, manchmal gibt es auch eine Englischstunde, wo man einfach das Buch aufschlägt und sagt: „So, das machen wir jetzt hier.“ Und es gibt natürlich

auch Themen, die einem selber mehr Spaß machen und mehr- ja, also, Englischlehrwerke haben finde ich oft auch nicht unbedingt ein Ziel, wo ich mich manchmal frage: „Ja, OK, was machen wir jetzt mit dem Wissen, das wir hier erworben haben?“ Also die Stunden gibt es natürlich auch. Von daher stimme ich nicht voll zu, sondern eher zu.

I: Bitte beschreiben Sie einmal, wie Sie das den Schüler*innen dann klar machen, in so einem Fall, wo Sie versuchen Relevanz herzustellen?

L2: Zum Beispiel, wenn es um Bestellen von Essen geht oder so. Dass man einfach sagt, wenn man eben in England mal unterwegs ist - da hat man auch immer Kinder, die mal in England waren -, dass man dann sagt: „So, wenn du jetzt dastehst, versteht dich keiner, wenn du auf Deutsch antwortest. Das heißt, du musst schon versuchen, eben auf Englisch dann bestellen zu können. Und das üben wir dann eben hier.“ So etwas zum Beispiel. (Transkript Lehrerin 2, Item 19)

Die Antwort von Lehrerin 2 wurde global mit der Kategorie „Vollständige Passung“ kodiert, da mehrere ihrer Unterkategorien in dem Interviewausschnitt vertreten sind. Es wird deutlich, dass Lehrerin 2 Bezug auf den Englischunterricht nimmt (1d; „Englischstunde“, „Englischlehrwerke“, „England“). Im ersten Teil der Antwort dokumentiert sich in Übereinstimmung mit dem Item die Ich-Perspektive, bevor Lehrerin 2 in Reaktion auf die Nachfrage zu einem verallgemeinernden „man“ übergeht (1c). Das im zweiten Teil der Antwort angeführte Beispiel entspricht der intendierten Iteminterpretation, da Lehrerin 2 durch Kontextualisierung der Aufgabe Relevanz sowie eine Verknüpfung zur Lebenswelt ihrer Schüler*innen herstellt (1a). Schließlich erscheint auch die gewählte Antwort („Stimme eher zu.“) vor dem Hintergrund der geschilderten Variabilität des Lehrkräfteverhaltens („Also die Stunden gibt es natürlich auch.“) schlüssig (1e).

Bei weiteren neun Items weisen die Interpretationen der Lehrerinnen hingegen ausschließlich keine und/oder nur eine eingeschränkte Passung mit der intendierten Iteminterpretation auf. Es ist auffällig, dass von diesen Items gleich vier dem Bereich corrective feedback im Englischunterricht entstammen.

Daraus ergibt sich ein breites Mittelfeld aus 26 Items, bei denen die Iteminterpretation durch die Lehrerinnen unterschiedlich ausfiel. Der Anteil an Interviews, in denen eine vollständige Passung zwischen intendierter Interpretation und der Iteminterpretation der Lehrerinnen vorlag, variiert zwischen einem Viertel (z. B. [meta-linguistic_cue]) und drei Vierteln (z. B. [simulation]). Es wurde lediglich ein Drittel der Fragebogenitems von den teilnehmenden Englischlehrerinnen ausschließlich im intendierten Sinn bearbeitet (s. Abb. 2).

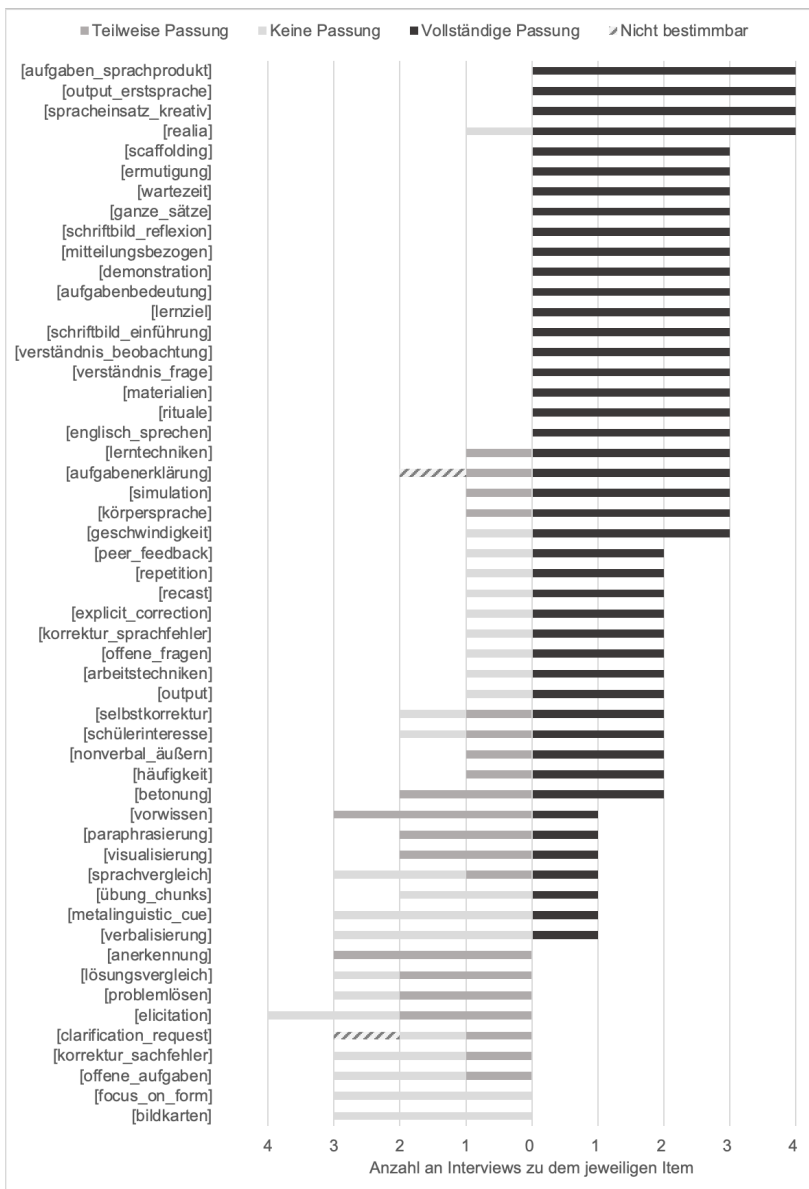


Abbildung 2: Passung zwischen dokumentierter und intendierter Interpretation der Lehrkräfteeitems

Im Falle des Schülerfragebogens zeichnet sich in der Übersicht der Globalurteile der 49 Items (s. Abb. 3, Anhang E) ein noch drastischeres Bild ab. In den Interviewdaten dokumentiert sich lediglich bei zehn Items eine vollständige Passung zu der intendierten Iteminterpretation über alle Interviews hinweg. Anders als bei dem Lehrkräftefragebogen bezieht sich ein Großteil dieser Items auf die durch die Englischlehrkraft eingesetzten Strategien zum corrective feedback oder Maßnahmen, die der Produktion von zielsprachlichem language output durch die Schüler*innen dienlich sind.

Bei 13 Items lassen die Schüleräußerungen aus den kognitiven Interviews ausschließlich auf keine und/oder nur eine eingeschränkte Passung zwischen intendierter und dokumentierter Iteminterpretation schließen. Dieser Anteil entspricht knapp einem Viertel des Schülerfragebogens. Obwohl die 13 Items alle Inhaltsbereiche des Fragebogens abbilden, beschreibt der Großteil der Items Maßnahmen, die einer besonders tiefen Informationsverarbeitung dienlich sind. Auch hierzu wird die Kodierung anhand eines Interviewausschnitts von Schüler*in 19 zu dem Item [lernetchniken] („Mein*e Englischlehrer*in bringt mir Tipps bei, wie ich besser für den Englischunterricht lernen kann (z. B. neue Wörter auf Englisch).“) beispielhaft veranschaulicht:

S19: Ja, stimme voll zu, weil der- wir haben gestern einen Englischtest geschrieben und da hatten manche halt Fehler drin bei „it's“ und „is“ und da hat die uns erklärt, wie man das richtig aufschreiben soll auf Englisch.

I: Gibt es in dem Satz ein Wort, das du nicht verstehst?

S19: [S19 liest das Item erneut.] Nö, ich kenne alle.

I: Kannst du den Satz bitte noch einmal in deinen eigenen Worten sagen?

S19: Ja, dass sie uns Tipps beibringt für den Englischunterricht und halt- und auch beibringt auf Englischunterricht was.

(Transkript Schüler*in 19, Item 29)

Aus der Schülerantwort geht hervor, dass die Fehlerbesprechung im Rahmen der Rückgabe eines Englischtests als Lerntipp eingeordnet wird. Diese Einordnung widerspricht jedoch der intendierten Iteminterpretation, die auf die Vermittlung metakognitiver Strategien zum selbstregulierten Lernen abzielt (3b). Auch die von Schüler*in 19 gewählte Wir-Perspektive ist aufgrund der Unterrichtssituation im Plenum zwar nachvollziehbar, jedoch nicht in Übereinstimmung mit der Perspektive aus dem Item (3c). Aus dem Interview lässt sich ableiten, dass das Wort „Tipps“ – vermutlich aufgrund seiner inhaltlichen Unschärfe – ungeeignet ist, um Schüler*innen zu metakognitiven Strategien zu befragen.

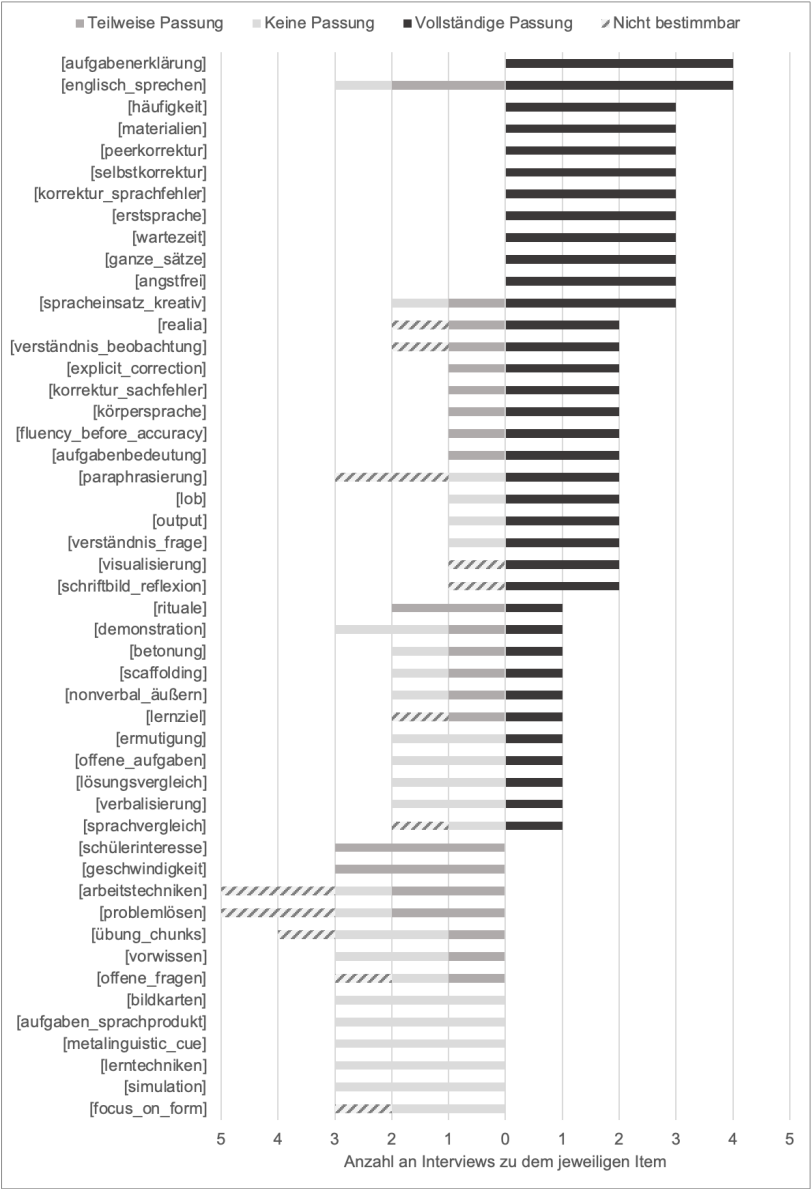


Abbildung 3: Passung zwischen dokumentierter und intendierter Interpretation der Schüleritems

Die Urteile über die Passung der Iteminterpretation fallen bei 26 Schüleritems uneindeutig aus. Der Anteil an vollständiger Passung variiert unter diesen Items zwischen einem Viertel (z. B. [demonstration]) und zwei Dritteln (z. B. [lob]). Weiter gilt für die Schüleritems, dass eine geringere sprachliche Komplexität, operationalisiert durch die Wortanzahl pro Item, mit einem valideren Itemverständnis einhergeht. Der Median der Itemlänge aller Items beträgt 18. Der Median der Itemlänge aller Items, die ausschließlich im intendierten Sinne interpretiert wurden, liegt bei 16; dagegen liegt er für die übrigen Items bei 20.

Aufgrund der hohen Diskrepanz zwischen intendierter und dokumentierter Iteminterpretation wurden alle Items, die keine vollständige Passung über alle Interviews hinweg aufweisen, kritisch geprüft und überarbeitet oder eliminiert. Dies betraf insgesamt 35 der 53 Lehrkräfteitems und 39 der 49 Schüleritems. Die verbliebenen Items wurden unverändert beibehalten (s. Anhang F).

5.2 Strategien zum Umgang mit häufigen Fehlertypen

Zur Beantwortung der Forschungsfragen 2a und 2b werden in diesem Abschnitt die häufigsten Fehlertypen, die in der Interviewstudie bei der Itembearbeitung aufkamen, zusammengefasst und es werden beispielhaft Strategien aufgezeigt, wie diesen Fehlertypen in der Itemüberarbeitung begegnet werden kann.

In den Schüleritems war eine nicht intendierte Iteminterpretation am häufigsten auf einen zu hohen Abstraktionsgrad in der Itemformulierung zurückzuführen. Dazu sei auf das Item [lerntechniken] verwiesen, das bereits in Abschnitt 5.1 diskutiert wurde. In seiner ursprünglichen Formulierung wies das Item keinen Bezug zu einer konkreten Unterrichtssituation oder metakognitiven Strategien auf. Dies führte dazu, dass die Schüler*innen das Wort „Tipps“ in den Interviews mit individuellen Vorstellungen füllten. Nach der Überarbeitung lautet das Item: „Mein*e Englischlehrer*in fragt mich am Ende der Englischstunde, was ich neu gelernt habe.“ In dieser Formulierung fokussiert das Item eine für Schüler*innen identifizierbare Unterrichtssituation, die jedoch offen genug gestaltet ist, um unterschiedliche Implementationen (im Sinne eines Oberflächenmerkmals von Unterricht) einzuschließen. In den Unterrichtshospitationen konnte beispielsweise beobachtet werden, dass manche Englischlehrerinnen eine mündliche Abfrage durchführten während andere digitale exit tickets¹ bearbeiten ließen. Zudem ist mit der Reflexion über die eigene Lernprogression eine konkrete kognitive Aktivität beschrieben.

Des Weiteren bildete die wörtliche Interpretation und Überbetonung von Itembestandteilen einen häufigen Fehlertyp seitens der Schüler*innen. Dies sei anhand des

¹ Bei dieser Aktivität beantworten Schüler*innen über eine digitale Anwendung schriftliche Impulse zur Reflexion ihres Lernfortschritts (z. B. State one thing you take away from today's class).

Items [vorwissen] („Mein*e Englischlehrer*in spricht mit mir über Sachen, die ich aus dem Alltag von zuhause kenne.“) verdeutlicht. In allen Interviews wurde der Ausdruck „von zuhause“ fokussiert, der zu einer Einschränkung auf den Wohnbereich („aus dem Haus“) und eigene Besitztümer („unsere Sachen“) führte, wie das Interview von Schüler*in 8 veranschaulicht:

S8: Stimme eher nicht zu. Weil im Alltag aus dem Haus haben wir quasi noch nie gemacht und wir haben halt auch noch nie gesprochen, was wir so im Haus haben.

I: Worüber sprichst du denn im Englischunterricht?

S8: Wir haben zum Beispiel über wie die Möbel im Haus heißen drüber gesprochen aber nicht über unsere Sachen auf Englisch haben wir gemacht. Dann haben wir noch über Farben und Zahlen auf Englisch gesprochen. (Transkript Schüler*in 8, Item 20)

Schüler*in 8 stimmt dem Item eher nicht zu, obwohl im Laufe des Interviews durchaus Inhaltsbereiche („Möbel“, „Farben und Zahlen“) aufgezählt werden, die der Alltagswelt der Schüler*innen zuzuordnen sind. Zudem deuteten weitere Schüleräußerungen darauf hin, dass das Item uneindeutig ist. Das Item kann entweder in Bezug auf das Vorwissen der Schüler*innen („Sachen, die ich kenne“) oder in Bezug auf den Lebensweltbezug des Unterrichtsgegenstands („Sachen aus dem Alltag von zuhause“) interpretiert werden. Diese Problematik wurde auch in den Lehrerinneninterviews mehrfach anhand des analog formulierten Lehrkräfteitems diskutiert. In der Überarbeitung wurde das Item [vorwissen] in zwei Items [vorwissen] („Mein*e Englischlehrer*in spricht mit mir über Sachen, die ich schon weiß.“) und [alltagserfahrung] („Mein*e Englischlehrer*in spricht mit mir über Sachen aus meinem Alltag.“) zerlegt.

In den Lehrkräfteinterviews stellte sich heraus, dass einige Items fremdsprachendidaktische Fachbegriffe enthielten, die den Englischlehrerinnen gar nicht oder nur teilweise bekannt waren. Dies galt für das Konzept der Lernaufgabe aus dem task-based language teaching und für alle Items der Skala zu corrective feedback. Die reine Umschreibung unterschiedlicher Strategien des corrective feedback führte bspw. dazu, dass Englischlehrerinnen dieselbe Strategie mit mehreren Items assoziierten, die jedoch verschiedene Strategien zum Inhalt hatten. Zur genaueren Differenzierung wurden die Items zu corrective feedback um Beispiele ergänzt. Dieses Vorgehen soll anhand des überarbeiteten Items [metalinguistic_cue] demonstriert werden: "Ich biete Hinweise zur Umformulierung an, die jedoch nicht die konkrete Fehlerkorrektur enthalten. – Schüler*in: I'd like to have three apple, please. – Lehrkraft: Think about the word 'apple' again. I'd like to have three...?" Während das Item zuvor nur das Bereitstellen metasprachlicher Hinweise als Feedbackstrategie beschrieb, enthält es nun einen Beispieldialog, in dem die Lehrkraft die Aufmerksamkeit der Schülerin auf das fehlerhafte Wort in ihrer Äußerung fokussiert.

6 Diskussion

Zum Nachweis der Inhaltsvalidität neu konstruierter Items zur fachspezifischen Erfassung kognitiver Aktivierung im Englischunterricht der Grundschule wurden kognitive Interviews mit Grundschüler*innen und Englischlehrerinnen geführt. Mittels strukturierender qualitativer Inhaltsanalyse wurde herausgearbeitet, dass bei einem Großteil der Items (35 aus 53 Lehrkräfteitems, 39 aus 49 Schüleritems) die Itemverarbeitung in mindestens einem Interview nicht der in der Entwicklung intendierten Iteminterpretation entsprach. Diese Befunde unterstreichen grundsätzlich die Notwendigkeit der systematischen Validierung bei der Konstruktion neuer Skalen; das gilt für jeden Forschungsgegenstand. Gleichzeitig zeigen die Ergebnisse eine Limitation zurückliegender empirischer Untersuchungen von Unterrichtsqualität auf, deren Instrumente gar nicht oder nur eingeschränkt validiert wurden. Die Studienergebnisse lassen sich zudem widerspruchsfrei in den bisherigen Forschungsdiskurs einordnen. Sie bestätigen die hohe Fehlerquote von Grundschüler*innen bei der Iteminterpretation (Lenske, 2016) zur Beurteilung kognitiver Aktivierung (Lenske & Praetorius, 2020) für den Englischunterricht.

In methodischer Hinsicht ist hervorzuheben, dass die vier in Abschnitt 2.1 beschriebenen Subskalen kognitiver Aktivierung über beide Probandengruppen hinweg unterschiedlich valide interpretiert werden. Items zum Einsatz von corrective feedback wurden von den Schüler*innen deutlich häufiger im intendierten Sinne interpretiert als von den Englischlehrerinnen. Dies könnte sich beispielsweise durch Itemperspektive und -referenz (Fauth et al., 2020) erklären lassen. Englischlehrkräfte reagieren während des Unterrichts auf eine Vielzahl von Fehlern unterschiedlicher Schüler*innen (Lyster & Ranta, 1997). Innerhalb von Sekunden entscheiden sie nicht nur, ob sie den Fehler thematisieren, sondern auch, wie sie ihn im Falle einer Korrektur implizit (z. B. mittels sprachlich korrekter Umformulierung) oder explizit (z. B. durch einen direkten Hinweis auf den Fehler) korrigieren. Damit müssen Englischlehrkräfte diese vielfältigen Interaktionen bei Items zu corrective feedback in Form einer Selbstbeurteilung zu einem Globalurteil akkumulieren. Schüler*innen hingegen können das Item aus einer Mischung von Selbst- und Fremdbeurteilung beantworten, indem sie die deutlich geringere Anzahl an Interaktionen mit ihrer Englischlehrkraft reflektieren. Um dieser Herausforderung seitens der Lehrkräfte zu begegnen, wurden die Items der Subskala corrective feedback während der Überarbeitung gemäß unterschiedlicher Feedbackstrategien differenziert und durch Beispiele veranschaulicht. Die Fragebogenpilotierung wird zeigen, ob diese Maßnahmen ausreichend sind, oder ob einige Aspekte des Englischunterrichts aufgrund ihrer Komplexität möglicherweise nicht aus allen Perspektiven valide beurteilt werden können.

Zusammenfassend seien die Items mit besonders hoher Passung in der Iteminterpretation charakterisiert. Im Falle der Lehrkräfte wurden besonders solche Items

mehrheitlich im intendierten Sinn verstanden, die zuvor im Rahmen einer Online-Expert*innenbefragung (Guttke & Porsch, im Druck) von Lehrkräften und Seminarleitungen als prototypisch für den Englischunterricht benannt wurden. Es ist möglich, dass diese Items valide bearbeitet werden, da ihre Inhalte aufgrund einer wiederholten Thematisierung in der Lehrkräfteausbildung (Lehramtsstudium und Vorbereitungsdienst) oder überdurchschnittlich häufigen Auftretens im eigenen Unterricht besonders salient für die Lehrkräfte sind. Im Falle der Schüler*innen zeichnen sich Items mit mehrheitlich vollständiger Passung dadurch aus, dass sie besonders handlungsnah formuliert sind. Die Ergebnisse der beiden Probandengruppen deuten an, dass die Zusammenarbeit mit Akteur*innen der Schulpraxis unerlässlich für die Konstruktion inhaltsvalider Items im untersuchten Themenbereich ist.

Abschließend sei auf die Limitationen der berichteten Studie hingewiesen. Bei den Proband*innen handelte es sich um eine Gelegenheitsstichprobe. Die Stichprobe umfasst Grundschulen unterschiedlicher Sozialindexstufen und Schüler*innen mit unterschiedlichen Erstsprachen, sie ist in Bezug auf diese Merkmale dennoch eingeschränkt repräsentativ. Obwohl Cohen's κ zu den restriktiveren Maßen der Interrater-Übereinstimmung zählt, deutet der Wert von .58 auf eine unzureichende Übereinstimmung in der Kodierung der Lehrkräfteinterviews hin. Rückblickend ist dieser geringe Wert auf die Mittelkategorie (Eingeschränkte Passung) zurückzuführen, da die Identifikation einzelner von den Proband*innen zur Begründung und Illustration von Itemantworten herangezogener Indikatoren hoch-inferent ausfiel. Zudem wurden die überarbeiteten Items keiner erneuten Prüfung unterzogen. Dadurch bleibt offen, inwiefern die Itemüberarbeitung zu einer Verbesserung des Verständnisses geführt hat. Im Rahmen der Pilotierung wird dazu das Itemverständnis der Schüler*innen über eine optionale „Verstehe ich nicht“-Option erfasst.

Digitaler Anhang

Der Anhang ist nicht Teil der Publikation. Die Verantwortung für den Inhalt des Anhangs liegt bei den darin ausgewiesenen Autor*innen: <https://www.doi.org/10.17185/dupublico/81375>

Literatur

- Bromme, R. (2014). *Der Lehrer als Experte*. Münster: Waxmann.
- Caspar, F. & Wirtz, M. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*. Münster: Waxmann.
- Collins, D. (2015). *Cognitive Interviewing Practice*. Los Angeles: SAGE. <https://www.doi.org/10.4135/9781473910102>
- De Jong, R. & Westerhof, K. (2001). The quality of student ratings of teacher behavior. *Learning Environments Research*, 4, 51-85. <https://www.doi.org/10.1023/A:1011402608575>

- Emde, M. & Fuchs, M. (2012). Datenqualität in standardisierten Interviews mit Kindern. In F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung* (S. 335-349). Weinheim: Beltz Juventa.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014a). Grundsschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive: Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28 (3), 127-137. <https://www.doi.org/10.1024/1010-0652/a000129>
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014b). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1-9. <https://www.doi.org/10.1016/j.learninstruc.2013.07.001>
- Fauth, B., Göllner, R., Lense, G., Praetorius, A.-K. & Wagner, W. (2020). Who sees what? Conceptual considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik, Beiheft*, 138-155.
- Göllner, R., Wagner, W., Klieme, E., Lüdtke, O., Nagengast, B. & Trautwein, U. (2016). Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven. In Bundesministerium für Bildung und Forschung (Hrsg.), *Forschungsvorhaben in Anknüpfung an Large-Scale-Assessments* (S. 63-82). Bundesministerium für Bildung und Forschung.
- Guttke, J. (2023). Kognitive Aktivierung im Fremdsprachenunterricht: Ein systematisches Review von Forschungsarbeiten aus dem deutschsprachigen Raum. *Zeitschrift für Fremdsprachenforschung*, 34 (2), 145-176.
- Guttke, J. & Porsch, R. (im Druck). Inhaltsvalidierung von Fragebogenitems zu kognitiver Aktivierung im Englischunterricht der Grundschule mittels Expert:innenbefragung. In V. Lohe, A. Lindl & P. Kirchhoff (Hrsg.), *Unterrichtsqualität in schulischen Fremdsprachen*. Münster: Waxmann.
- Jansen, N., Decristan, J. & Fauth, B. (2022). Individuelle Nutzung unterrichtlicher Angebote – Zur Bedeutung von Lernvoraussetzungen und Unterrichtsbeteiligung. *Unterrichtswissenschaft*, 50, 157-183. <https://www.doi.org/10.1007/s42010-021-00141-8>
- Kersten, K. (2021). L2 input and characteristics of instructional techniques in early foreign language classrooms. *The European Journal of Applied Linguistics and TEFL*, 10 (2), 27-59.
- Klieme, E. (2022). Unterrichtsqualität. In M. Haring, C. Rohlf & M. Gläser-Zikuda (Hrsg.), *Handbuch Schulpädagogik* (S. 411-426). Stuttgart: UTB.
- Kolb, A. & Schocker, M. (2021). *Teaching English in the Primary School*. Stuttgart: Klett.
- Krippendorff, K. (2019). *Content analysis: An introduction to its methodology*. Los Angeles: SAGE. <https://www.doi.org/10.4135/9781071878781>
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9 (3), 231-251. <https://www.doi.org/10.1007/s10984-006-9015-7>
- Lense, G. (2016). *Schülerfeedback in der Grundschule: Untersuchungen zur Validität*. Münster: Waxmann.
- Lense, G. & Praetorius, A.-K. (2020). Schülerfeedback – was steckt hinter dem Kreuz auf dem Fragebogen? *Empirische Pädagogik*, 34 (1), 11-29.
- Lipowsky, F. (2020). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 69-118). Berlin: Springer. https://www.doi.org/10.1007/978-3-662-61403-7_4
- Lipowsky, F. & Bleck, V. (2019). Was wissen wir über guten Unterricht? – Ein Update. In U. Steffens & R. Messner (Hrsg.), *Unterrichtsqualität: Konzepte und Bilanzen gelingenden Lehrens und Lernens* (S. 219-250). Münster: Waxmann.
- Lipowsky, F., Drollinger-Vetter, B., Klieme, E., Pauli, C. & Reusser, K. (2018). Generische und fachdidaktische Dimensionen von Unterrichtsqualität – Zwei Seiten einer Medaille? In M. Martens, K. Rabenstein, K. Bräu, M. Fetzner, H. Gresch, I. Hardy & C. Schelle (Hrsg.), *Konstruktionen von Fachlichkeit* (S. 183-202). Bad Heilbrunn: Verlag Julius Klinkhardt.
- Loewen, S. (2020). *Introduction to instructed second language acquisition*. London: Routledge. <https://www.doi.org/10.4324/9781315616797>
- Lyster, R. & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19 (1), 37-66. <https://www.doi.org/10.1017/S0272263197001034>

- Mayring, P. (2022). Qualitative Inhaltsanalyse: Grundlagen und Techniken. Weinheim: Beltz Juventa. https://www.doi.org/10.1007/978-3-658-37985-8_43
- Porsch, R. & Wilden, E. (2022). Teaching English out-of-field in primary school – Differences in professional characteristics and effects on self-assessed instructional quality. In L. Hobbs & R. Porsch (Hrsg.), *Out-of-field teaching across teaching disciplines and contexts* (S. 117-134). Berlin: Springer.
- Praetorius, A.-K., Rogh, W. & Kleickmann, T. (2020a). Blinde Flecken des Modells der drei Basisdimensionen von Unterrichtsqualität? Das Modell im Spiegel einer internationalen Synthese von Merkmalen der Unterrichtsqualität. *Unterrichtswissenschaft*, 48, 303-318. <https://www.doi.org/10.1007/s42010-020-00072-w>
- Praetorius, A.-K., Herrmann, C., Gerlach, E., Zülsdorf-Kersting, M., Heinitz, B. & Nehring, A. (2020b). Unterrichtsqualität in den Fachdidaktiken im deutschsprachigen Raum – zwischen Generik und Fachspezifik. *Unterrichtswissenschaft*, 48 (3), 409-446. <https://www.doi.org/10.1007/s42010-020-00082-8>
- Rieser, S. & Decristan, J. (2023). Kognitive Aktivierung in Befragungen von Schülerinnen und Schülern. *Zeitschrift für Pädagogische Psychologie*, online first. <https://www.doi.org/10.1024/1010-0652/a000359>
- Röhl, S. & Rollet, W. (2020). Alles nur die Nettigkeit der Lehrkraft? Die Communion der Lehrkräfte als Erklärung für den Halo-Bias in Schülerbefragungen zur Unterrichtsqualität. *Empirische Pädagogik*, 34 (1), 30-45.
- Tourangeau, R., Rips, L. J. & Rasinski, K. (2012). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungs-Modelle der Wirkweise des Unterrichts. *Zeitschrift für Pädagogik, Beiheft*, 63-80.
- Wafudu, S. J., Kamin, Y. B. & Marcel, D. (2022). Validity and reliability of a questionnaire developed to explore quality assurance components for teaching and learning in vocational and technical education. *Humanities and Social Sciences Communication*, 9, 303. <https://www.doi.org/10.1057/s41599-022-01306-1>
- Wilden, E. (2021). Fachspezifische Aspekte von Unterrichtsqualität im Schulfach Englisch. *Unterrichtswissenschaft*, 49 (2), 211-219. <https://www.doi.org/10.1007/s42010-021-00105-y>
- Willis, G. (2005). *Cognitive interviewing*. Los Angeles: SAGE. <https://www.doi.org/10.4135/9781412983655>
- Wisniewski, B., Röhl, S. & Fauth, B. (2022). The perception problem: A comparison of teachers' self-perceptions and students' perceptions of instructional quality. *Learning Environments Research*, 25, 775-802. <https://www.doi.org/10.1007/s10984-021-09397-4>

Autor*innen

Joel Guttke, M. Ed., Universität Duisburg-Essen, Institut für Anglophone Studien, Didaktik des Englischen

Prof. Dr. Raphaela Porsch, Otto-von-Guericke-Universität Magdeburg, Institut I: Bildung, Beruf und Medien, Fachbereich Erziehungswissenschaft

Korrespondenz an: joel.guttke@uni-due.de

Rezension

Wolfgang Klug

Melanie Misamer (2023): Machtsensibilität in der Sozialen Arbeit: Grundwissen für reflektiertes Handeln

1 Die Autorin

Dr. Melanie Misamer ist seit 2021 Professorin für Methoden und Konzepte Sozialer Arbeit in der Gesundheitsförderung an der HAWK Göttingen, Fakultät für Ingenieurwissenschaften und Gesundheit, Bereich: Gesundheitscampus. Nach ihrem Studium der Sozialen Arbeit schloss sie 2018 ihre Promotion im Bereich der Pädagogischen Psychologie ab (Thema: Macht und Machtmittel in der Schule). Sie forscht und publiziert im Bereich der Themen Machtsensibilität, multidisziplinäres dimensionales Machtkonzept, Machtanwendung und (ethische) Prinzipien, Gesundheit und Macht, Gerechtigkeit und Vertrauen (<https://melaniemisamer.de/>).

2 Inhalt

Nach einer Einführung befasst sich das Werk im ersten Kapitel mit der Frage *Um was geht es bei Macht?* Hier werden verschiedene Definitionen von Macht und Machtmitteln kurz vorgestellt (z. B. Weber, Machiavelli, Luhmann, u. a.), schließlich Macht als „neutrales Potenzial“ und als „Vermögen, auf das Verhalten anderer Einfluss zu nehmen“ (Misamer, 2023, S. 23) definiert. Deutlich herausgearbeitet werden das Thema Machtanwendung und die Folgen von Machtnutzung. Anhand klassischer Experimente (Milgram-Experiment; Stanford-Experiment) eröffnet die Autorin das Kernthema: Gefahren der Machtanwendung durch Korruption und die Folgen davon. Basierend auf den empirischen Erkenntnissen von Dacher Keltner, der im Folgenden noch häufig zitiert wird, formuliert sie die zentrale These: „Mit dem Innehaben von Macht geht jedoch allmählich der Bezug zu anderen verloren“ (Misamer, 2023, S. 51). Aus dieser ist deutlich herauszuhören, dass die Autorin von einer konstruktiven (im Einklang mit ethischen Standards stehenden) und einer destruktiven (Adressaten und Adressatinnen der Sozialen Arbeit schadenden) Nutzung von Macht ausgeht.

Mit dem in der Sozialen Arbeit bekannten Phänomen der „Ohnmacht“ beschäftigt sich das 2. Kapitel. Ohnmacht – als Erleben von Hilflosigkeit angesichts erlebter

Machtverhältnisse – habe Auswirkungen auf die körperliche und psychische Gesundheit. Als Erklärungsmodell wird die Theorie der „Erlernten Hilflosigkeit“ von Seligman herangezogen. *Ohnmächtig-Machen in der Praxis* lautet ein Kapitel, das mit Beispielen auf der systemischen Ebene, der Gruppenebene und der individuellen Ebene aufwartet. Unter anderem geht es um Macht von Bestrafungs- und Belohnungssystemen und den Folgen der Expertenmacht. Die Autorin warnt vor Machtmissbrauch in Form von destruktiven Varianten der Übergriffigkeit. Diesen Gefahren ließe sich mittels Reflexion begegnen.

Das 3. Kapitel ist dem Kernthema „Machtsensibilität“ gewidmet. Diese „beschreibt eine konstitutive Empfindsamkeit gegenüber Machtanwendung“ (Misamer, 2023, S. 97), die u. a. das Wissen um Korruptionsmechanismen und die „Fallstricke der eigenen Wahrnehmung“ (ebd.) beinhaltet. Für das Verständnis der Autorin zu ihrem Thema ist die anthropologische Aussage: Der „Wille zur Macht [ist] ein menschlicher Wesenszug. Jeder Mensch strebt mehr oder weniger nach Macht und das ist ein Tabuthema“ (Misamer, 2023, S. 98). Im Sinne derachtsensibilisierung müsse dieses Tabu „gebrochen“ werden, „um eine Basis für eine Sensibilisierung für die eigene Machtnutzung zu schaffen“ (ebd.). Dass es mit derachtsensibilität in helfenden Berufen nicht allzu weit her ist, zeigt die Autorin anhand von zwei eigenen empirischen Studien, die sie an der Schule und mit Sozialarbeitenden durchgeführt hat: Bei beiden Berufsgruppen stellt sich heraus, dass die einzelne Berufsfachkraft ihr eigenes Verhalten konstruktiver einschätzt als das Verhalten der gesamten Berufsgruppe. Gegenmaßnahmen zu Machtmissbrauch könnten nach Ansicht der Autorin die Etablierung klarer berufsethischer Prinzipien sein. Betont wird in diesem Zusammenhang auch immer wieder die Notwendigkeit von Selbstreflexion. Thematisiert werden anschließend die positiven Auswirkungen vonachtsensibilität (z. B. in der Vorbildfunktion für Klientinnen und Klienten oder höherer Handlungssicherheit).

„Wie lässt sich Macht in der Sozialen Arbeit konstruktiv nutzen?“, fragt die Autorin schließlich im vorletzten Kapitel, in dem es um die praktischen Folgerungen geht. Herzstück ist dabei ein Abschnitt, der mit „Eine kleine Methodensammlung: Partizipative Strategien in der Praxis“ überschrieben ist. Zu finden ist ein „konstruktives Handwerkszeug für den Methodenkoffer der Arbeitspraxis“ (Misamer, 2023, S. 149ff.).

Im 5. und letzten Kapitel werden Erlebnisse aus der Praxis zum Umgang mit Macht in der Arbeitspraxis geschildert.

3 Diskussion

Unter dreierlei Gesichtspunkten soll das Buch von Melanie Misamer diskutiert werden:

- ▶ 1. Anthropologisch-philosophisch
- ▶ 2. Sozialarbeitspraktisch
- ▶ 3. Gestaltung und Didaktik

Die Autorin zeigt neben ihrem sozialarbeiterisch-pädagogischen ein anthropologisch-philosophisches Interesse. Nicht nur, dass sehr viele Philosophen (wenn auch kurz) angesprochen und zitiert werden, es nehmen weiterhin philosophisch bedeutsame Konstrukte einen bedeutenden Platz ein. So werden beispielsweise die Themen „Gerechtigkeit“ und gewisse anthropologische Positionen (z. B. der Mensch als Träger des „Willens zur Macht“) an prominenter Stelle diskutiert. Zu den aus Sicht des Rezensenten überzeugendsten und innovativsten Teilen des Buches gehört die Verschränkung von Gerechtigkeit, Vertrauen und konstruktiver Machtanwendung als Aspekte sozialarbeiterischer Beziehungsgestaltung. Hier lohnt es sich, weiter nachzudenken und diese philosophisch inspirierten Diskurse empirisch zu untermauern.

Dass sich die Autorin auf eine Eingangsdefinition von Macht als „neutrales Potenzial“ und „Vermögen, auf das Verhalten anderer Einfluss zu nehmen“ festlegt, ist schon deshalb nicht zu beanstanden, weil diese Definition im Großen und Ganzen mit der weit verbreiteten Definition von Max Weber übereinstimmt. Einer weiteren Festlegung, dass Macht „grundsätzlich nicht an Personen, sondern an den Status [...] oder die Situation [...]“ (Misamer, 2023, S. 20) gebunden ist, bringt zwar die das Buch bestimmende Grundthese der asymmetrischen Machtbeziehung zwischen Fachkraft und deren Klientel voran, führt aber auch in eine Dichotomie von Fachkraft als mächtiger versus Klient*in als ohnmächtiger Person. Diese Dichotomie ist es denn auch, die andere faktisch in der Praxis vorkommende Machtkonstellationen nicht mehr in den Blick treten lässt. So arbeitet Zobrist in seiner empirischen Untersuchung der Machtdynamiken in Arbeitsbeziehungen zwischen Sozialarbeiter*innen und Jugendlichen heraus, dass Jugendliche selbst im Zwangskontext und in anscheinend aussichtslos unterlegener Stellung ihre Machtpotenziale entfalten. Ein befragter Sozialarbeiter meint resigniert: „letztendlich hat der Jugendliche finde ich/ hockt immer noch am längeren Hebel, weil eben wenn er irgendwo hingeht und sagt ‚hey er MACHT das nicht‘ oder er SIEHT es überhaupt nicht, da wird das/ wird das kein Erfolg werden“ (Zobrist, 2023, S. 312). Insbesondere, wenn Sozialarbeitende die Grundkonzeption ihres Wirkens als „Koproduktion“ zwischen sich und ihrer Klientel konstruieren oder gar die Klienten*innen als einzigen „Produzenten“ sehen (wie Schaarschuch & Oelerich, 2005), wird aus der asymmetrischen, scheinbar so einseitigen Machtbeziehung, sehr schnell eine für die Fachkraft prekäre Machtbasis. Vielleicht würde eine stärkere Differenzierung zwischen Macht und Gewalt, wie Hannah Arendt (1970) sie vorschlägt, oder die interaktionell-kommunikative Zugangsweise von Niklas Luhmann (1975/2012) dazu beitragen, Macht nicht einseitig als Positi-

onsmerkmal, sondern als Zuschreibungsphänomen zu begreifen, das entsteht, wenn Menschen (egal, in welcher Position) einen anderen ermächtigen, indem sie ihm Macht über sich verleihen.

Das Buch ist aus sozialarbeitspraktischer Sicht sehr lesenswert. Zum einen trifft es ein Desiderat, denn in der Tat ist das Thema „Macht“ in weiten Kreisen des Berufsstandes, der sich das „Helfen“ auf die Fahnen schreibt, ein Tabu. Ob man es „durchbrechen“ kann, noch dazu mithilfe der Berufsethik eines Berufsverbandes, dem nur ein Bruchteil der Sozialarbeitenden angehört, ist allerdings eine andere Frage. Sehr hilfreich sind die zahllosen Methoden und Anregungen zu einer machtsensiblen Vorgehensweise, die das Buch, besonders im hinteren Teil, auch für Praktiker*innen sehr wertvoll machen. Es ist ein lohnendes Unterfangen, gerade in der Sozialen Arbeit darauf hinzuweisen, wie notwendig die Überprüfung der eigenen Vorannahmen (z. B. von „beratungsresistenten“ Klienten) wäre, wie dies die Autorin unter machtheoretischen Gesichtspunkten zwingend aufzeigt. Der Rezensent kann dies aus einem eigenen aktuellen Forschungsprojekt um die Notwendigkeit und Schwierigkeit von Reflexion in der Praxis (Klug & Mirus, 2024) nur bestätigen. Angesichts der Praxisnähe des Buches bleibt es ein wenig erstaunlich, warum die Soziale Arbeit im Zwangskontext etwas stiefmütterlich behandelt wird. Die bei Misamer für diesen Sozialarbeitszweig ohnedies wenig verwendete Literatur (Dollinger 2010, Lutz, Huber und Schierz 2013, Heuer 2018) betrachtet Zwangskontexte allesamt von einem sozialpädagogisch-kritischen Standpunkt aus und bietet einen – aus Sicht des Rezensenten – nicht wirklich erhellenden Zugang. Besser wäre sicher das Einbeziehen von theoretisch wie praktisch inspirierenden Werken, z. B. Zobrist und Kähler (2017) oder Trotter (2014).

Ein besonderes Lob verdient die gestalterisch-didaktische Aufbereitung des Themas. Die Autorin entwickelt ein ausgesprochenes Geschick, den Leser und die Leserin an die Hand zu nehmen und durch nicht immer ganz leichte Theoreme zu führen. Dabei versäumt sie es nie, ihre Theorien mit einer Vielzahl von empirischen Belegen abzusichern und diese gerade im sozialpsychologischen Teil unterhaltsam mit einschlägigen Studien zu präsentieren. Der Spagat zwischen einer solchermaßen theoriebezogenen, aber gleichzeitig auf die Verbesserung der Praxis zielenden Bearbeitung ist gelungen. Vielleicht hätten die am Ende etwas „verlorenen“ Praxiszitate noch besser eingearbeitet werden können, aber das ist dann doch Geschmacksache.

Wenn Studierende ein Referat gut finden, sagen sie bisweilen: „Es wurde nie langweilig!“. Voila: Beim Lesen wurde es dem Rezensenten nie langweilig.

Vollständige bibliografische Angabe des rezensierten Werkes:

Misamer, M. (2023). *Machtsensibilität in der Sozialen Arbeit. Grundwissen für reflektiertes Handeln* (1. Aufl.). Stuttgart: Kohlhammer. <https://www.doi.org/10.17433/978-3-17-042186-8>

Literatur

- Arendt, H. (1970). Macht und Gewalt. München: Piper.
- Klug, W. & Mirus, G. (2024). Wie reflektieren Fachkräfte der Sozialen Arbeit ihre Arbeitsbeziehungen mit Klienten/innen? (Forschungsprojekt). Katholische Universität Eichstätt-Ingolstadt.
- Luhmann, N. (1975/2012). Macht (4. Aufl.). Konstanz: UVK-Verlagsgesellschaft.
- Schaarschuch, A. & Oelerich, G. (2005). Theoretische Grundlagen und Perspektiven sozialpädagogischer Nutzerforschung (S. 9-25). In G. Oelerich & A. Schaarschuch (Hrsg.), Soziale Dienstleistungen aus Nutzersicht. Zum Gebrauchswert Sozialer Arbeit. München: Ernst Reinhardt-Verlag.
- Trotter, C. (2014). Working with involuntary clients: A guide to practice (3rd ed.). Abingdon: Routledge. <https://www.doi.org/10.4324/9781315880587>
- Zobrist, P. (2023). Machtdynamiken in Arbeitsbeziehungen zwischen Sozialarbeiter*innen und Jugendlichen im Zwangskontext der Jugendanwaltschaft (Dissertation). Zürich: Universität. Verfügbar unter: https://www.zora.uzh.ch/id/eprint/238071_122.02.2024.
- Zobrist, P. & Kähler, H. D. (2017). Soziale Arbeit in Zwangskontexten. Wie unerwünschte Hilfe erfolgreich sein kann (3. Aufl.). München: Ernst Reinhardt-Verlag.

Rezensent

Prof. Dr. Wolfgang Klug, Katholische Universität Eichstätt-Ingolstadt, Fakultät für Soziale Arbeit

Korrespondenz an: wolfgang.klug@ku.de

Impressum

Empirische Pädagogik

Zeitschrift zu Theorie und Praxis erziehungswissenschaftlicher Forschung

ISSN 0931-5020

Verlag Empirische Pädagogik

Bürgerstraße 23, 76829 Landau/Pfalz

Telefon: +49 6341 280 32180; Telefax: +49 6341 280 32166

E-Mail: ep@vep-landau.de

Homepage: www.vep-landau.de

Beirat

Roland Arbinger (Landau), Susanne Buch (Wuppertal), Monika Buhl (Heidelberg), Marten Clausen (Duisburg), Andreas Frey (Mannheim), Julia Fluck (Aachen), Michaela Gläser-Zikuda (Erlangen-Nürnberg), Reinhold S. Jäger (Landau), Gisela Kammermeyer (Landau), Katharina Maag Merki (Zürich), Jörn Sparfeldt (Saarbrücken), Olga Zlatkin-Troitschanskaia (Mainz).

Redaktion

Gabriele E. Dlugosch, Ingmar Hosenfeld, Heike Liepelt, Tamara Nold, Sarah Poersch, Michael Zimmer-Müller.

Beiträge

Die Empirische Pädagogik veröffentlicht Beiträge zu folgenden Rubriken: Originalarbeiten, Forschungsmethoden, Forum, Historische Seite und Rezensionen. Beiträge nimmt die Redaktion entgegen.

Manuskriptgestaltung

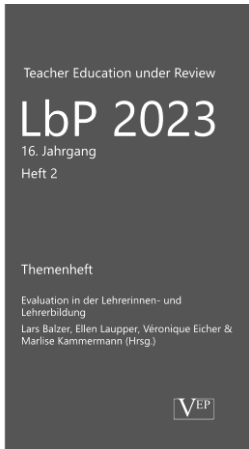
Die aktuellen Autorenrichtlinien finden Sie auf unserer Webseite unter www.vep-landau.de/publizieren/. Abgabe des Manuskripts per E-Mail als Word- oder pdf-Datei an: ep@vep-landau.de.

Open Access

Die Konditionen zur Publikation als Open Access finden Sie auf unserer Webseite unter www.vep-landau.de/publizieren/.

Erscheinungsweise/Preis

Die Zeitschrift erscheint viermal jährlich. Der Bezugspreis beträgt € 60.00/Jahr zzgl. Porto. Kündigung bis spätestens 8 Wochen vor Jahresende.



Lehrerbildung auf dem Prüfstand Teacher Education under Review

Beiträge zur Lehrerbildung mit Hilfe empirischer Untersuchungen, Analysen und Diskussionen

ISSN 1867-2779

Zwei Hefte pro Jahr, € 40.00/Jahr (zzgl. Versandkosten); Einzelhefte ab € 15.90;

Digitale Versionen sind als DRM-freies PDF über den Shop (<https://www.vep-landau.de>) erhältlich

Wir bieten die Möglichkeit, einzelne Beiträge oder ganze Hefte als Open Access zu publizieren. Details unter www.vep-landau.de/publizieren/

Letzte Themenhefte Jahrgang 16 (2023)

- ▶ Balzer, L., Laupper, E., Eicher, V. & Kammermann, M. (Hrsg.). (2023). Evaluation in der Lehrerinnen- und Lehrerbildung. (Lehrerbildung auf dem Prüfstand, 16 (2), Themenheft). Landau: Verlag Empirische Pädagogik. ISBN 978-3-944996-95-0. <https://doi.org/10.62350/GHBH7658> **OPEN ACCESS**

Jahrgang 15 (2022)

- ▶ Schaper, N. & Vogelsang, C. (Hrsg.). (2022). Erfassung und Förderung von Reflexionskompetenz in der Lehrerbildung. (Lehrerbildung auf dem Prüfstand, 15 (1), Themenheft). Landau: Verlag Empirische Pädagogik. ISBN 978-3-944996-86-8. **OPEN ACCESS**

Jahrgang 14 (2021)

- ▶ Stadler-Altman, U. (Hrsg.). (2021). Digitale Lehrerbildung. (Lehrerbildung auf dem Prüfstand, 14 (1), Themenheft). Landau: Verlag Empirische Pädagogik. ISBN 978-3-944996-77-6. **OPEN ACCESS**

Editorial

- Roland Happ, Kristina Kögler, Jacqueline Schmidt & Marc Egloffstein: Lehren und Lernen mit und über Künstliche Intelligenz in der Aus- und Weiterbildung

Thementeil

- Jacqueline Schmidt & Roland Happ: Validierung eines Tests zur Erfassung des Grundlagenwissens zu Künstlicher Intelligenz von angehenden Lehrkräften
- Jule Hangen & Eveline Wuttke: Künstliche Intelligenz verstehen lernen – Ergebnisse eines Trainingsprogramms zur Förderung von Startup-Fähigkeiten im Bereich KI
- Sabine Seufert, Lukas Spirgi, Jan Delcker, Joana Heil & Dirk Ifenthaler: Umgang mit KI-Robotern: maschinelle Übersetzer, Textgeneratoren, Chatbots & Co – Eine empirische Studie bei Erstsemester Studierenden
- Josef Guggemos, Jacqueline Schmidt & Roland Happ: Eine Frage von Macht: Einstellungen angehender Lehrkräfte zu den ethischen Grundsätzen des Einsatzes künstlicher Intelligenz in der Bildung
- Marc Egloffstein, Kristina Kögler & Dirk Ifenthaler: Evidenzgestützte Entwicklung von onlinebasierten Lernangeboten zu Künstlicher Intelligenz in der beruflichen Bildung: Stakeholder-Perspektiven und Implementierung
- Julia Pargmann, Anna Leube, Florian Berding, Elisabeth Riebenbauer, Karin Rebmann, Andreas Slopinski & Michael Gillert: Die elektronisch-didaktische Assistenz (EDDA) zur Unterrichtsplanung auf Basis künstlicher Intelligenz: Funktionsweise, Anwendungsbereiche und Forschungsperspektiven

Allgemeiner Teil

- Joel Guttke & Raphaela Porsch: Kognitive Interviews als Methode der Inhaltsvalidierung von Fragebogenitems zur Erfassung kognitiver Aktivierung im Englischunterricht der Grundschule

Rezension

- Wolfgang Klug: Melanie Misamer (2023): Machtsensibilität in der Sozialen Arbeit: Grundwissen für reflektiertes Handeln



www.vep-landau.de

ISSN 0931-5020

ISBN 978-3-944996-97-4



9 783944 996974