

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Von der Quelle zur KI
ChatGPT im modernen Geschichtsunterricht

verfasst von | submitted by

Andreas Lukas Blauensteiner BEd

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Education (MEd)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 199 502 511 02

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Lehramt Sek (AB) Unterrichtsfach
Biologie und Umweltbildung Unterrichtsfach
Geschichte und Politische Bildung

Betreut von | Supervisor:

Mag. Mag. Dr. Alexander Preisinger

Abstract

Künstliche Intelligenzen, insbesondere Modelle wie ChatGPT, werden von Schüler:innen zunehmend als Hilfe bei der Bearbeitung schulischer Aufgaben herangezogen. Es ist jedoch unklar, wie sich der Einsatz dieser Tools auf kompetenzorientiertes, historisches Lernen auswirkt und welche Herangehensweisen dabei zu beobachten sind. Ziel der vorliegenden Arbeit ist es zu klären, welche Strategien Schüler:innen beim Umgang mit ChatGPT zeigen, wenn sie kompetenzorientierte, in den Anforderungsbereichen I–III verortete Aufgaben im Geschichtsunterricht bearbeiten, und wie diese Strategien mit Einstellungen, Selbstwirksamkeit und Vertrauen der Schüler:innen gegenüber ChatGPT sowie mit deren Nutzungsmustern korrelieren.

Um dies zu beantworten, wurde eine explorative Mixed-Methods-Studie mit 14 Schüler:innen des WIFI St. Pölten durchgeführt. Die teilnehmenden Personen bearbeiteten dabei in Paaren drei Aufgaben mittels ChatGPT. Die Bearbeitung wurde im Rahmen der „Think-Aloud“-Methode per Screen- und Audioaufnahme aufgezeichnet, transkribiert und in MaxQDA mit einem dreidimensionalen Kodiersystem (KI-Strategien, FUER-Kompetenzen, Intensitätsniveau) ausgewertet. Zusätzlich erfasste ein standardisierter Online-Fragebogen Skalen zu Wissen, Nutzung, Einstellungen, Selbstwirksamkeit und Vertrauen gegenüber ChatGPT, die deskriptiv und mit Spearman-Korrelationen in SPSS analysiert wurden.

Aus der qualitativen Analyse ergaben sich die drei Strategietypen „reflexiv-selektive“, „outputorientierte“ und „situativ-explorative“ KI-Nutzung. Historische Kompetenzen konnten vor allem dort gemessen werden, wo ChatGPT gezielt begrenzt und durch KI-freie Denkphasen ergänzt wurde. Eine hohe Nutzungsfrequenz und hohe Selbstwirksamkeit gingen in der analysierten Stichprobe dagegen eher mit outputorientierter, delegierender Nutzung einher. Der situativ-explorative Typ schwankte schließlich zwischen beiden Modi, ohne einen stabilen Bearbeitungsstil auszubilden. Die Ergebnisse zeigen, dass nicht die Intensität der KI-Nutzung, sondern deren didaktische Einbettung entscheidend für die Förderung kompetenzorientierten, historischen Denkens ist, woraus sich Implikationen für die Unterrichtsgestaltung sowie die Lehrer:innenbildung ergeben.

1.	Einleitung	1
2.	Grundlagen und Funktionsweise von ChatGPT	3
2.1.	Begriffsdefinitionen.....	3
2.2.	Training großer Sprachmodelle (LLM).....	5
2.3.	Funktionsweise von ChatGPT	7
2.4.	Grenzen von LLMs	7
3.	Rahmenbedingungen für den Einsatz von Künstlicher Intelligenz.....	9
3.1.	KI-Regulierung auf EU-Ebene – Der AI-Act.....	9
3.2.	KI-Strategie in Österreich – AIMAT 2030	11
3.3.	KI in Schule und Unterricht	12
3.3.1.	Systematik, Chancen und Risiken.....	12
3.3.2.	KI im österreichischen Bildungssystem	15
3.4.	Forschungsstand zum Einsatz von ChatGPT in Bildung und Schule.....	16
4.	ChatGPT in der Geschichtsdidaktik	20
4.1.	FUER-Kompetenzen	20
4.2.	Digitale Medien im Geschichtsunterricht.....	23
4.3.	ChatGPT im Geschichtsunterricht.....	28
5.	Methodik	33
5.1.	Forschungsdesign und Forschungsfragen	33
5.2.	Stichprobe, Setting und Durchführung der Datenerhebung	34
5.3.	Erhebungsinstrumente.....	36
5.3.1.	Qualitative Instrumente	36
5.3.2.	Quantitative Instrumente	40
5.4.	Datenaufbereitung und Skalenbildung	43
5.5.	Strategiekategorien und Korrelationsanalyse	45
5.6.	Theoretische Grundlagen der Kategorienbildung	46
5.7.	Entwicklung und finales Kodiersystem.....	47
5.8.	Codebuch.....	49

6.	Qualitative Ergebnisse.....	61
6.1.	Deskriptive Verteilung der Codes.....	61
6.2.	Grundlegende Muster im qualitativen Material	63
6.3.	Typenbildung nach Kuckartz im Umgang mit ChatGPT	64
6.3.1.	Merkmalsraum und Typenbildung	65
6.3.2.	Beschreibung der Strategietypen und Zuordnung der Fälle.....	68
7.	Quantitative Ergebnisse.....	72
7.1.	Deskriptive Statistik.....	72
7.2.	Korrelationstabelle der erhobenen Skalen.....	74
7.3.	Explorative Zusammenhänge zwischen Strategietypen und Skalen	76
8.	Diskussion.....	78
8.1.	Strategietypen im Kontext der quantitativen Skalen.....	78
8.2.	Reflexion des Forschungsdesigns	79
9.	Fazit.....	81
9.1.	Beantwortung der Forschungsfragen.....	81
9.2.	Implikationen für Unterricht und Lehrer:innenbildung.....	83
9.2.1.	Didaktische Chancen und Risiken von ChatGPT im Geschichtsunterricht	83
9.2.2.	Leitlinien für kompetenzorientierten KI-Einsatz und Lehrer:innenbildung.....	84
9.3.	Limitationen	86
9.4.	Ausblick.....	87
10.	Literaturverzeichnis.....	89
11.	Appendix	96
11.1.	Qualitativer Fragebogen.....	96
11.2.	Quantitativer Fragebogen.....	101

1. Einleitung

Als Stephen Hawking 2018 starb, hinterließ er der Welt nicht nur bahnbrechende physikalische Erkenntnisse, sondern auch eine eindringliche Warnung vor den Folgen von Künstlicher Intelligenz:

„It will either be the best thing that’s ever happened to us, or it will be the worst thing. If we’re not careful, it very well may be the last thing.“¹

Ähnlich äußerte sich auch Sam Altman, Geschäftsführer von OpenAI, jener Firma, welche ChatGPT entwickelte. Er beschrieb dabei 2015 KI einerseits zwar als Motor für wirtschaftliche Innovation, andererseits aber auch als potenzielle Bedrohung für die Menschheit.²

Trotz dieser Warnungen ist der Einsatz von Künstlicher Intelligenz in den vergangenen Jahren in zahlreiche Lebensbereiche vorgedrungen. In Unternehmen werden KI-Systeme im Rennen um mehr Effizienz zur Automatisierung des Supports, zur Erstellung personalisierter Werbung und zur Optimierung interner Verwaltungsprozesse eingesetzt.³ Gleichzeitig nutzen Privatpersonen Künstliche Intelligenz, um sich ihren Alltag und ihre Arbeit zu erleichtern, beispielsweise beim Organisieren digitaler Dokumente, beim Verfassen von Texten oder für simulierte Streitgespräche mit Chatbots, um andere Perspektiven und Gegenargumente auszuloten.⁴

Auch das Bildungssystem ist von dieser „KI-Revolution“ betroffen. Eine Studie des Österreichischen Bundesverlags ÖBV zeigt beispielsweise, dass KI-Tools bereits in erheblichem Ausmaß an Schulen verwendet werden.⁵ Damit stellt sich nicht mehr die Frage, ob Bildungsinstitutionen auf KI reagieren sollen, sondern vielmehr wie sie dies tun.

Als Lehrperson bin ich selbst unmittelbar von KI-Systemen wie ChatGPT betroffen. Einerseits nutzen meine Schüler:innen das Modell, um sich die Erledigung ihrer Aufgaben zu erleichtern, andererseits setze ich es selbst ein, um meine Unterrichtsvorbereitung effizienter zu gestalten.

¹ Abigail Higgins, Stephen Hawking’s final warning for humanity: AI is coming for us. In: Vox, Future Perfect (16.10.2018), <https://www.vox.com/future-perfect/2018/10/16/17978596/stephen-hawking-ai-climate-change-robots-future-universe-earth> (15.4.2025).

² Charles I. Jones, At What Point Do We Decide AI’s Risks Outweigh Its Promise? In: Stanford Graduate School of Business, Insights (29.5.2024), <https://www.gsb.stanford.edu/insights/what-point-do-we-decide-ais-risks-outweigh-its-promise> (15.4.2025).

³ Jim Holdsworth, Matthew Finio, AI Examples & Business Use Cases. In: IBM Think, <https://www.ibm.com/think/topics/artificial-intelligence-business-use-cases> (15.4.2025).

⁴ Francesca Paris, Larry Buchanan, 35 Ways Real People Are Using A.I. Right Now. In: The New York Times, The Upshot (14.4.2023), <https://www.nytimes.com/interactive/2023/04/14/upshot/up-ai-uses.html> (15.4.2025).

⁵ öbv, Wie verbreitet ist Künstliche Intelligenz an Schulen? In: öbv Magazin (27.11.2023), <https://www.oebv.at/magazin/wie-verbreitet-ist-kuenstliche-intelligenz-an-schulen> (15.4.2025).

Außerdem beschäftige ich mich im privaten Bereich eingehend mit Python, einer wichtigen Programmiersprache im KI-Bereich. In diesem Sinne ist es mir ein persönliches Anliegen, die Digitalisierung und Forschung bezüglich Künstlicher Intelligenz im Bildungssektor voranzutreiben, damit Schulen und Lehrpersonen auf das Aufkommen dieser neuen Technologien zeitgerecht reagieren können und die Chancen, die diese bieten, nicht „verschlafen“.

Die vorliegende Arbeit beschäftigt sich mit dem Einsatz von ChatGPT im Geschichtsunterricht. Da viele Schüler:innen den Chatbot bereits jetzt ohne Supervision in vielen Situationen nutzen, um Aufgaben zu erledigen oder zu erleichtern, erscheint es wenig sinnvoll, das Tool zu ignorieren oder zu verbieten. Stattdessen stellt sich die Frage, wie ChatGPT so in den Unterricht integriert werden kann, dass damit historisches Lernen im Sinne kompetenzorientierter Modelle ermöglicht wird.

Vor diesem Hintergrund lautet die zentrale Forschungsfrage der Arbeit:

„Welche Strategien lassen sich beim Umgang mit ChatGPT identifizieren, die Schüler:innen nutzen, um kompetenzorientierte, in den Anforderungsbereichen I–III verortete geschichtsdidaktische Aufgaben zu bearbeiten?“

Ergänzend wird untersucht, inwiefern diese Strategien mit quantitativen Merkmalen wie Einstellungen und Wissen bezüglich KI zusammenhängen. Dabei ist das Ziel, erste empirische Einsichten in die Rolle von KI-Nutzung für historisches Lernen zu gewinnen und daraus Implikationen für den kompetenzorientierten Einsatz von ChatGPT im Geschichtsunterricht abzuleiten.

Im anschließenden Theorieteil werden zentrale Begriffe und technische Grundlagen von KI und „Large Language Models“ erläutert, der bildungswissenschaftliche und geschichtsdidaktische Rahmen skizziert sowie der aktuelle Forschungsstand zu ChatGPT im Bildungskontext dargestellt. Anschließend wird die Methodik, die Ergebnisse und deren Diskussion erläutert sowie Implikationen für Unterricht und weitere Forschung gegeben.

2. Grundlagen und Funktionsweise von ChatGPT

Die vorliegende Arbeit nutzt als Kernelement zur Datensammlung ChatGPT („Generative Pre-Trained Transformer“). Im folgenden Kapitel werden dessen Grundlagen und die wichtigsten Definitionen besprochen. Da die dahinterliegende Informatik jedoch komplex ist, wird nur ein kurzer Überblick gegeben, welcher für das Verstehen der Arbeit als ausreichend erachtet wird. Als Hauptquellen dienen dabei ein Artikel aus dem Blog „Digital Humanities am Deutschen Historischen Institut Paris (DHIP)“ von Mareike König sowie die Website des Herstellers OpenAI selbst.

2.1. Begriffsdefinitionen

Für das Verständnis der weiteren Arbeit ist es notwendig, zentrale Begriffe des Themenfeldes zu klären und zu definieren. Im Folgenden werden daher „Künstliche Intelligenz“, „maschinelles Lernen“ sowie „generative KI“ auf Basis von GPT-Modellen umrissen.

Die Basis von „Chatbots“ wie ChatGPT bilden sogenannte „Large Language Models“ (LLM), die Sprache auf statistischer Basis modellieren. Ihr besonderes Potenzial ergibt sich aus der Verarbeitung großer Textmengen, wodurch sie in der Lage sind, Sprache zu analysieren, nachzuvollziehen und selbst zu produzieren. Da solche Systeme eigenständig Inhalte wie Texte, Bilder, Code oder audiovisuelle Medien hervorbringen können, werden sie unter dem Begriff „generative künstliche Intelligenz“ („GenAI“) zusammengefasst.⁶ Neben ChatGPT gibt es weitere große Modelle auf dem Markt, welche von verschiedenen Tech-Firmen betreut werden, beispielsweise Google, Microsoft oder Meta.⁷ Für den schulischen Kontext ist dabei insbesondere relevant, dass all diese Systeme in natürlicher Sprache bedient werden können und die Basisversionen kostenlos sind, womit sie für Schüler:innen gut zugänglich sind.

Das Bundesministerium für Bildung, Wissenschaft und Forschung (BMBWF) definiert „Künstliche Intelligenz“ in seiner „Strategie der Bundesregierung für Künstliche Intelligenz“ (2021) als Computersysteme, die „intelligentes Verhalten“ zeigen, indem sie Aufgaben ausführen, für die früher menschliche Entscheidungen und Intelligenz notwendig waren. Solche Systeme analysieren ihre Umgebung und treffen auf dieser Basis selbstständig Entscheidungen,

⁶ Mareike König, ChatGPT und Co. in den Geschichtswissenschaften – Grundlagen, Prompts und Praxisbeispiele. In: Digital Humanities am DHIP (19.8.2024), <https://dhdhi.hypotheses.org/9197> (25.6.2025); OpenAI, GPT-4, <https://openai.com/de-DE/index/gpt-4-research/> (25.6.2025).

⁷ Steffen Albrecht, ChatGPT und andere Computermodele zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen (TAB-Hintergrundpapier 26, Berlin 2023), doi.org/10.5445/IR/1000158070.

um bestimmte Ziele zu erreichen.⁸ Die Europäische Union beschreibt KI im „Artificial Intelligence Act“ (2024) ergänzend als maschinengestützte Systeme, die in unterschiedlichem Ausmaß autonom agieren, Eingaben verarbeiten und daraus Ausgaben ableiten, die ihre Umgebungen beeinflussen können.⁹ Beide Definitionen machen deutlich, dass es sich bei KI weniger um eine einzelne Technologie als eher um ein breites Forschungsfeld handelt, welches seit der Mitte des 20. Jahrhunderts entwickelt wurde und in den letzten Jahren durch neue Durchbrüche stark an gesellschaftlicher Bedeutung gewonnen hat.¹⁰

Als besonders aktuelles Teilgebiet der KI gilt das maschinelle Lernen. In einer Handreichung zu KI in der Bildung beschreibt das BMBWF maschinelles Lernen als Verfahren, bei dem Computersysteme aus Daten und Erfahrungen „lernen“ und dadurch ihre Leistung bei der Bearbeitung bestimmter Aufgaben verbessern können. Dies reicht von einfachen statistischen Modellen bis hin zu künstlichen neuronalen Netzen. Das Lernen kann dabei unter Vorgabe gewünschter Ergebnisse oder ohne vorab definierte Zielgrößen autonom erfolgen.¹¹

Die vorliegende Arbeit fokussiert innerhalb des KI-Feldes auf „generative Künstliche Intelligenz“, insbesondere auf „Generative Pre-Trained Transformer“-Modelle. Dabei handelt es sich um Sprachmodelle, die mit einem sehr umfangreichen Textkorpus trainiert wurden und in der Lage sind, aus Texteingaben völlig neue sprachliche Ausgaben zu erzeugen. Das bekannteste Beispiel für diese „Familie“ von Modellen ist „ChatGPT“ der Firma OpenAI. Der Chatbot kann auf Basis der im Training verwendeten Daten eigenständig Texte generieren, Fragen beantworten und in einen Dialog mit Nutzer:innen treten, wodurch der Eindruck eines menschlichen Gegenübers entsteht.¹²

Aufbauend auf diesen begrifflichen Grundlagen wird im folgenden Abschnitt die Funktionsweise generativer Sprachmodelle, insbesondere von GPT-basierten Systemen wie ChatGPT, erläutert, wobei der Fokus auf bildungsrelevanten Aspekten liegt.

⁸ Bundesministerium für Bildung, Wissenschaft und Forschung, Strategie der Bundesregierung für Künstliche Intelligenz. Artificial Intelligence Mission Austria 2030 (Wien 2021), 16, <https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/strategie-bundesregierung.html> (25.4.2025).

⁹ Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Verordnung über künstliche Intelligenz), Art. 12, <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (25.4.2025).

¹⁰ Bundesministerium für Bildung, Wissenschaft und Forschung, Auseinandersetzung mit Künstlicher Intelligenz im Bildungssystem (Wien 2023), 3-6, https://www.bmbwf.gv.at/Themen/schule/zrp/ki/ki_asbs.html (25.4.2025).

¹¹ BMBWF, Auseinandersetzung mit Künstlicher Intelligenz im Bildungssystem, 7.

¹² Simone Grassini, Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. In: Education Sciences 13 (2023) 692, doi.org/10.3390/educsci13070692.

2.2. Training großer Sprachmodelle (LLM)

OpenAI selbst gibt auf seiner Seite an, dass GPT-4 darauf trainiert wurde, „das nächste Wort in einem Dokument vorherzusagen“.¹³ In dieser Hinsicht kann es, vereinfacht gesagt, wie eine komplexe Wahrscheinlichkeitsrechnung betrachtet werden. Large Language Models wie die GPT-Modelle müssen allerdings vorher trainiert werden, um sinnvolle „Vorhersagungen“ zu liefern. Dabei werden zwei Phasen durchlaufen: Ein Pre-Training und anschließend eine Art Fine-Tuning der Ergebnisse.

Das Pre-Training von ChatGPT basiert auf der Verarbeitung im Voraus heruntergeladener und gespeicherter Internetdaten, die zur Schulung des zugrunde liegenden neuronalen Netzes genutzt werden. Für GPT-3 stammt dieser Datensatz dabei beispielsweise aus dem September 2021, was für die Beurteilung der Aktualität der gelieferten Ergebnisse durchaus eine Rolle spielen kann.¹⁴ Auch für das in der Arbeit verwendete Modell GPT-4 sieht das ähnlich aus, wie der Hersteller OpenAI auf seiner Seite dazu schreibt:

*„GPT-4 verfügt im Allgemeinen nicht über Kenntnisse zu Ereignissen, die nach dem Wissensstichtag (September 2021) eingetreten sind, und lernt nicht aus seinen Erfahrungen.“*¹⁵

Hierbei ist der genannte „Wissensstichtag“ wohl jener Tag, an dem der Datensatz aus dem Internet heruntergeladen und für das Training des Modells verwendet wurde. Allerdings muss angemerkt werden, dass das Modell GPT-4 Plus, welches für diese Arbeit verwendet wurde, in bestimmten Situationen sehr wohl Zugriff auf das Internet hat und somit auch auf Ereignisse nach diesem Stichtag Bezug nehmen kann. Den Kern des Modells stellen dabei weiterhin die Offline-Daten aus 2021 dar, jedoch kann GPT-4 Plus, wenn es die Gegebenheit als sinnvoll erachtet, auf Echtzeit-Daten aus dem Internet über eine Erweiterung mit Microsoft Bing zugreifen. Dabei formuliert das Modell einen eigenen Search-Prompt, welcher mit Bing verarbeitet wird. Die Suchergebnisse werden dann an GPT-4 Plus zurückgesendet und in die Ausgabe eingearbeitet.¹⁶

Bezüglich des Grunddatensatzes ist nicht im Detail bekannt, welche Teile des Internets im Speziellen einbezogen bzw. heruntergeladen wurden. Gesichert ist jedoch die Verwendung von Quellen wie dem „common crawl“, also öffentlich einsehbare Webseiten aller Art, Wikipedia,

¹³ OpenAI, GPT-4.

¹⁴ König, ChatGPT und Co.

¹⁵ OpenAI, GPT-4.

¹⁶ Xuan Ji, Does GPT-4 have access to the internet using AI? Unveiling the truth. In: BytePlus, <https://www.byteplus.com/en/topic/413088> (25.8.2025).

verschiedenen online verfügbaren Büchern, Diskussionsforen wie Reddit sowie Programmierplattformen. Diese Datensammlung wird vor dem Training transformiert und in kleinere Einheiten zerlegt, sogenannte „Token“, deren Kontexte vom Modell in einem vieldimensionalen Vektorraum abgebildet werden, wodurch sich statistische Beziehungen zwischen Wörtern und Phrasen ergeben, welche anschließend für die Textgenerierung genutzt werden können.¹⁷ Für den Unterricht ist insbesondere diese Kombination aus großem, aber intransparentem Datensatz und festem Wissensstichtag bedeutsam, da sie die Aktualität und Nachvollziehbarkeit der von ChatGPT generierten Inhalte stark einschränkt.

Der zweite Trainingsschritt eines Sprachmodells ist das sogenannte „Fine-Tuning“, welches insbesondere die Dialogfähigkeit hervorbringen soll. Es gibt verschiedene Arten, wie dieses feinere Anpassen vonstattengeht, jedoch nutzen die GPT-Modelle hierfür „Reinforcement Learning from Human Feedback“ (RLHF). Dafür werden eigens erstellte Datensätze eingesetzt, die aus menschlich formulierten Fragen und den korrekten Antworten dazu bestehen. Die generierten Antworten werden von Entwickler:innen überprüft und unerwünschtes Verhalten durch Anpassungen der Parameter zwischen den einzelnen Tokens reduziert. Im Gegensatz zum energie- und kostenintensiven Pre-Training erweist sich das Fine-Tuning als ressourcenschonender, weshalb es nicht nur einmalig, sondern kontinuierlich durchgeführt wird.¹⁸

OpenAI gibt außerdem an, dass der Prozess hinter diesem Feedback darauf basiert, dass die „Entscheidungen“ der KI sich zuerst zufällig aus dem für sie erstellten Netz ergeben und jedes menschliche Feedback eine Art „Belohnung“ darstellt, nach der sich die Parameter im Netz verändern. Mit zunehmender Verbesserung wird Feedback besonders wichtig bei Fällen, in denen die KI unsicher ist, welche Verhaltensweise besser ist. Dadurch verfeinern sich die Ergebnisse fortlaufend. Eine Herausforderung bleibt jedoch, dass der Ertrag dieses Prozesses von den Entwickler:innen abhängt. Diese müssen dafür sorgen, dass die gewählten Aufgaben als auch das gegebene Feedback qualitativ hochwertig ausfallen.¹⁹ Hierbei wird zusätzlich erwähnt, dass Feedback, wenn es nicht richtig eingesetzt wird, die Gesamtleistung des Modells sogar verringern kann.²⁰ Für den schulischen Einsatz bedeutet dies, dass auch Vorstellungen und Überzeugungen der Entwickler:innen bzw. diesbezügliche Anweisungen oder Vorgaben Einfluss auf die Antworten des Modells haben.

¹⁷ König, ChatGPT und Co.

¹⁸ König, ChatGPT und Co.

¹⁹ OpenAI, Learning from human preferences, <https://openai.com/index/learning-from-human-preferences/> (25.6.2025).

²⁰ OpenAI, GPT-4.

2.3. Funktionsweise von ChatGPT

Wie bereits weiter oben selbst von OpenAI beschrieben, ist die grundlegende Funktionsweise der GPT-Modelle eine Vorhersage des statistisch am besten geeigneten nächsten Wortes. Auf Basis des vorangegangenen Pre-Trainings wählt das Modell als Reaktion auf einen Prompt eines Users Schritt für Schritt jene Wörter aus, die im jeweiligen Kontext am plausibelsten erscheinen.²¹

Hierbei ist anzumerken, dass das Modell nicht immer das statistisch wahrscheinlichste Wort nimmt. Würde es das tun, würde der entstehende Text sich sehr „flach“ lesen und von Nutzer:innen als „computerhaft“ wahrgenommen werden. Um diesen Umstand zu verhindern, werden manchmal auch weiter unten in der „Liste“ stehende Wörter genutzt, wodurch die Texte kreativer werden und der Wortlaut der ausgegebenen Antwort sich trotz ähnlicher oder sogar gleicher Inputs dennoch unterscheidet.²²

Wesentlich für die vorliegende Arbeit ist, dass LLMs, auch wenn es so wirkt, nicht in der Lage sind, die Frage oder die gegebene Antwort zu „verstehen“. ChatGPT „denkt“ nicht logisch über ein Problem oder eine Fragestellung nach. Es gibt keine internen rationalen „Denkprozesse“, wie sie bei einem Menschen normalerweise auftreten. Die Modelle sind nicht darauf ausgelegt, eine Lüge zu erkennen oder den Wahrheitsgehalt der eigenen Ausgaben zu beurteilen. Es liegt ihnen keine Moral zugrunde, anhand derer sie antworten, sondern nur ein Sprachmodell und dazugehörige Gewichtungen, mit deren Hilfe sie aus den Trainingsdaten das nächste „wahrscheinlichste“ Wort, welches passen könnte, vorhersagen können. Auch wenn durch das weiter oben beschriebene Fine-Tuning mittels menschlichen Feedbacks ungewolltes Verhalten verringert werden kann, kann man das System dennoch dazu verleiten, inhaltlich falsche oder problematische Inhalte zu produzieren.²³

2.4. Grenzen von LLMs

Die Grenzen von Modellen wie ChatGPT hängen mit ihrem internen Aufbau bzw. dem zugrundeliegenden Prinzip zusammen. Die erste große Aufgabe, welche sich Entwickler:innen bei der Erschaffung eines verlässlichen LLMs stellen müssen, ist es, einen qualitativ hochwertigen Grunddatensatz zu erstellen. Dieser beinhaltet meist zwar eine große Fülle an Daten aus verschiedensten Quellen, allerdings ist dabei keine Objektivität oder Verlässlichkeit gegeben. Falschinformationen, Verzerrungen und oberflächliche Behandlungen bestimmter

²¹ König, ChatGPT und Co.

²² König, ChatGPT und Co.

²³ OpenAI, GPT-4.

Themen können darin genauso vorkommen wie beispielsweise problematische Meinungen oder Vorurteile aus politischer und menschenrechtlicher Sicht. Dieses Problem kann, wie weiter oben beschrieben, durch Fine-Tuning zwar eingedämmt, jedoch nicht vollständig behoben werden. Weiters kann die Anbindung von GPT-4 an Echtzeit-Informationen via Bing dabei helfen, in Einzelfällen aktuellere Ergebnisse zu liefern, jedoch kann auch dieses Feature nicht gänzlich problematische Inhalte vermeiden.

Damit ist eines der größten Probleme, unter denen LLMs wie ChatGPT leiden, Verzerrungen („Bias“) des zugrundeliegenden Datensatzes. Darunter wird verstanden, dass der Stammdatensatz aufgrund seiner Subjektivität problematische politische oder moralische Züge aufweist, die das LLM in seiner Ausgabe beeinflussen. Zusätzlich ist das Internet keine ausgeglichene Datensammlung. Viele Themen sind überrepräsentiert, andere unterrepräsentiert. So dominiert der globale Norden bei der Produktion von Texten, welche für die GPT-Modelle eingesetzt werden, und es werden dementsprechende Vorurteile und Perspektiven stärker gewichtet. Auch Fine-Tuning kann diesen Bias nicht in einer zufriedenstellenden Weise ausgleichen, sondern führt in manchen Fällen sogar zu gegenteiligen, nicht wünschenswerten Ergebnissen.²⁴

Zusätzlich zu diesen Verzerrungen gibt es noch ein anderes schwerwiegendes Problem bei der Nutzung von LLMs, den „Halluzinationen“. Wie aufgezeigt, basiert der Funktionsmechanismus von ChatGPT auf einer statistischen Vorhersagewahrscheinlichkeit für das nächste Token bzw. Wort. Dadurch kann es auch zu frei erfundenen Ausgaben kommen. Ein Beispiel dafür ist die Quellen- oder Buchsuche mittels LLMs. ChatGPT erfindet dafür oft einfach Buchtitel oder auch die ISBN und gibt eine zufällige 13-stellige Zahl aus. Je spezifischeres Wissen abgefragt wird, desto höher die Wahrscheinlichkeit, dass das Modell anfängt, Sachverhalte zu „halluzinieren“. Die genaue Ursache für diesen Umstand ist Gegenstand der derzeitigen Forschung in diesem Bereich. Neue Modelle konnten gegenüber älteren die Rate, mit der diese frei erfundenen Ergebnisse auftreten, bereits reduzieren, jedoch ist das Problem noch immer vorhanden.²⁵

Für den Einsatz im Unterricht bedeutet dies, dass Schüler:innen lernen müssen, Antworten von LLMs nicht blind zu vertrauen, sondern diese im Sinne historischer Quellenkritik zu prüfen. ChatGPT kann eine Unterstützung sein, ist in seiner jetzigen Form aber weder neutral noch durchgängig verlässlich.

²⁴ König, ChatGPT und Co.

²⁵ König, ChatGPT und Co.

3. Rahmenbedingungen für den Einsatz von Künstlicher Intelligenz

3.1. KI-Regulierung auf EU-Ebene – Der AI-Act

In der öffentlichen Diskussion bezüglich KI wird seit einigen Jahren immer wieder darauf hingewiesen, dass Europa im internationalen Vergleich bei Investitionen in Künstliche Intelligenz hinter Ländern wie den USA und China zurückliegt.²⁶ Teilweise wird sogar davon gesprochen, Europa habe wesentliche Teile der „KI-Revolution“ und der Digitalisierung verpasst.²⁷ Betrachtet man ausschließlich das Investitionsvolumen, zeigen verfügbare Schätzungen tatsächlich deutlich geringere Ausgaben für KI-Forschung und -Anwendungen in der EU im Vergleich zu den USA (European Court of Auditors, 2024; Wirtschaftskammer Österreich, 2024), was in Abbildung 1 ersichtlich gemacht wird.²⁸

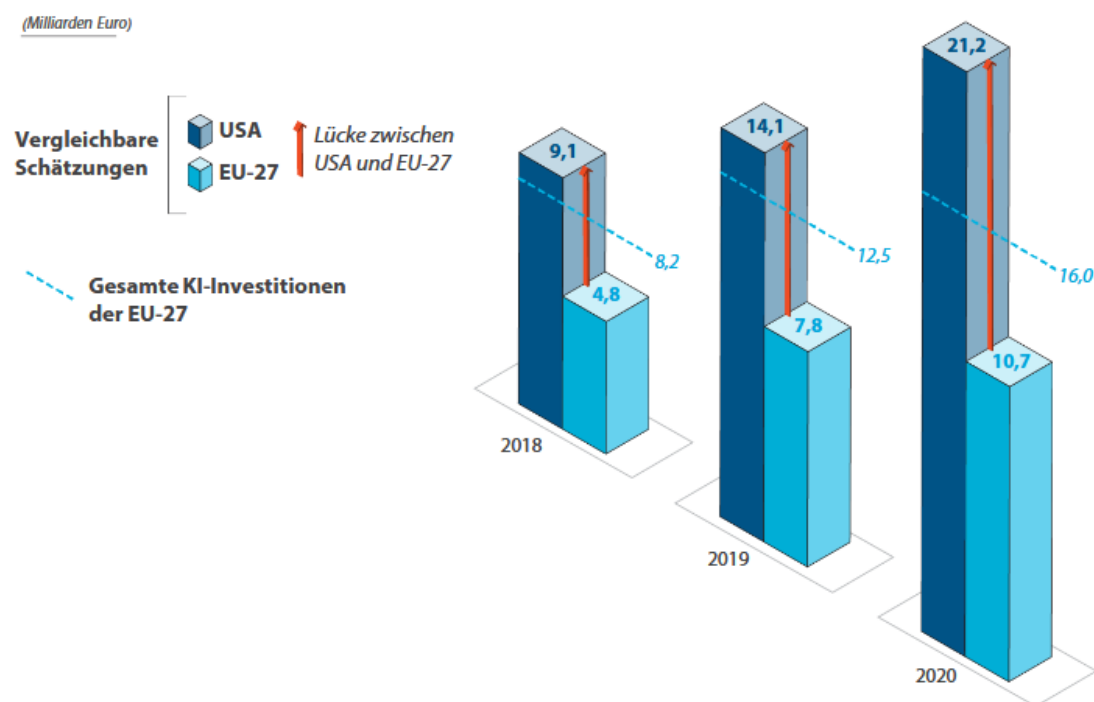


Abbildung 1: Gesamte KI-Investitionen der EU-27 und der USA 2018–2020 (in Mrd. Euro).
Quelle: Europäischer Rechnungshof, Die Ambitionen der EU im Bereich der künstlichen Intelligenz: Mehr Governance und verstärkte, gezielter ausgerichtete Investitionen sind zukunftsentscheidend (Sonderbericht 2/2024, Luxemburg 2024) 23.

²⁶ Daniel Castro, Michael McLaughlin, Eline Chivot, Who is winning the AI race: China, the EU or the United States? (Washington 2019), <https://www2.datainnovation.org/2019-china-eu-us-ai.pdf> (14.6.2025); Gregor Kramper, USA und China hängen Europa bei der Zukunftsindustrie Künstlicher Intelligenz ab. In: stern.de (3.2.2019), <https://www.stern.de/digital/technik/kuenstliche-intelligenz---usa-und-china-haengen-europa-ab-8563640.html> (14.6.2025).

²⁷ Martin Kugler, Europa hat die Digitalisierung verschlafen. In: Die Presse, Wort der Woche (14.9.2024), <https://www.diepresse.com/18861023/europa-hat-die-digitalisierung-verschlafen> (10.6.2025).

²⁸ Wirtschaftskammer Österreich, Europa: Verantwortungsvolle KI (25.6.2024), https://www.wko.at/epu/europa-verantwortungsvolle-ki#heading_Der_Nutzen_von_Regulierungen (10.6.2025); Europäischer Rechnungshof, Die Ambitionen der EU im Bereich der künstlichen Intelligenz, 23.

Unabhängig von der Höhe der Investitionen stellt sich jedoch die Frage, inwiefern der Einsatz von KI gesellschaftlich verantwortungsvoll und rechtlich abgesichert gestaltet werden kann. Vor diesem Hintergrund arbeitet die Europäische Union bereits seit mehreren Jahren an einem umfassenden Rechtsrahmen für KI.

Zentral ist hierbei der „Artificial Intelligence Act“ (2024). Er gilt als erstes umfassendes KI-Gesetz weltweit und verfolgt das Ziel, den vermehrten Einsatz von KI in einen transparenten, nicht-diskriminierenden und an europäischen Grundrechten orientierten Rechtsrahmen einzubetten. Kern des Gesetzes ist eine risikobasierte Einstufung von Künstlicher Intelligenz. KI-Systeme werden nach den potenziellen Gefahren, die von ihnen für Individuen und Gesellschaft ausgehen, in verschiedene Risikokategorien eingeteilt. Systeme mit „inakzeptablem Risiko“, etwa zur massenhaften biometrischen Überwachung oder zur manipulativen Beeinflussung besonders schutzbedürftiger Gruppen, sollen in der EU grundsätzlich verboten werden. Zusätzlich werden Hochrisiko-Bereiche, beispielsweise kritischer Infrastruktur, Strafverfolgung oder Justiz, mit strengen Auflagen versehen.

Die für diese Arbeit relevanten Teile der Verordnung betreffen die Transparenz der Nutzung von KI-Systemen. So werden Modelle wie ChatGPT nicht als „Hochrisiko-KI“ eingestuft, ihr Einsatz muss jedoch im Einklang mit dem EU-Urheberrecht stehen. Daraus ergeben sich Anforderungen an Autor:innen, die offenlegen müssen, wie und in welchem Ausmaß sie beim Kreieren ihrer Inhalte von KI unterstützt wurden. Die Systeme selbst müssen zudem dagegen abgesichert sein, dass durch sie illegale Materialien erzeugt werden. Zusätzlich müssen die verwendeten Trainingsdaten urheberrechtlich überprüft sowie in dieser Hinsicht gekennzeichnet sein. Dies betrifft besonders von KI erzeugte Bilder- oder Videodateien, um die Verbreitung von Fake News und Desinformation zu unterbinden. Einige Teile der Verordnung sind bereits in Kraft, so beispielsweise das Verbot von besonders risikoreichen KI-Systemen, wobei im Voraus eine eingehende Überprüfung und diesbezügliche Einordnung stattfinden muss. Andere Bestimmungen werden gestaffelt in den kommenden Jahren wirksam, wobei der AI-Act voraussichtlich ab Mitte 2026 in Kraft treten soll.²⁹ Damit bildet er den übergeordneten regulatorischen Rahmen, innerhalb dessen auch der Einsatz von ChatGPT im Bildungsbereich zu verorten ist.

²⁹ Verordnung (EU) 2024/1689, Verordnung über künstliche Intelligenz; Europäisches Parlament, KI-Gesetz: erste Regulierung der künstlichen Intelligenz (8.6.2023), <https://www.europarl.europa.eu/topics/de/article/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz> (10.6.2025).

3.2. KI-Strategie in Österreich – AIMAT 2030

Auf nationaler Ebene orientiert sich Österreich stark an diesem europäischen Rahmen und ist als Teil der EU in einer ähnlichen wirtschaftlichen Situation. Auch hier wird in der öffentlichen Debatte immer wieder darauf hingewiesen, dass das Land im internationalen Vergleich über ein relativ geringes Investitionsvolumen in KI-Forschung verfügt und damit Gefahr läuft, Talente und Innovationspotenzial an andere Staaten zu verlieren.³⁰

Als Antwort darauf wurde die „Artificial Intelligence Mission Austria 2030“ (AIMAT 2030) entwickelt. Diese nationale KI-Strategie verfolgt, ähnlich wie der AI-Act, eher einen wertorientierten Ansatz. Sie definiert dabei drei zentrale Zielrichtungen. Erstens die Nutzung von KI im Einklang mit Grund- und Menschenrechten, zweitens die Positionierung Österreichs als Forschungs- und Innovationsstandort im Bereich KI und drittens der Erhalt der Wettbewerbsfähigkeit der österreichischen Technologie- sowie Wirtschaftssektoren.³¹

Für den in dieser Arbeit wichtigen Bildungsbereich betont die AIMAT 2030 insbesondere die Bedeutung digitaler Kompetenzen beim Umgang mit Künstlicher Intelligenz. KI-bezogene Inhalte sollen dabei vermehrt in die Lehrer:innenbildung integriert werden. Zusätzlich sollen Quereinsteiger:innen aus den Bereichen der Wirtschaft und Industrie, welche über KI-Erfahrung aus dem Berufsleben verfügen, gezielt gefördert werden.³²

Gleichzeitig wird AIMAT 2030 in der aktuellen öffentlichen Diskussion bereits teilweise wieder als überholt kritisiert, da die Strategie vor dem Aufstieg großer Sprachmodelle wie ChatGPT entstand und diesbezügliche Entwicklungen nur unzureichend abbildet.³³ Unabhängig davon macht dieser Prozess deutlich, dass sich Österreich grundsätzlich zu einem wertorientierten, rechtskonformen Umgang mit KI bekennt und sich am „europäischen Weg“ orientiert. Für die vorliegende Arbeit bildet die AIMAT 2030 einen wichtigen bildungspolitischen Rahmen, innerhalb dessen der Einsatz von ChatGPT im Unterricht zu diskutieren ist.

³⁰ APA, Österreichs Unternehmen sind dabei den digitalen Wandel zu verschlafen. In: derStandard.at (13.2.2024), <https://www.derstandard.at/story/3000000207231/oesterreichs-unternehmen-sind-dabei-den-digitalen-wandel-zu-verschlafen> (10.6.2025); Wirtschaftskammer Österreich, Europa: Verantwortungsvolle KI; Martin *Stepanek*, Österreich verspielt seine KI-Zukunft. In: derStandard.at (15.6.2023), <https://www.derstandard.at/story/3000000174849/oesterreich-verspielt-seine-ki-zukunft> (10.6.2025).

³¹ BMBWF, Strategie der Bundesregierung für Künstliche Intelligenz, 9.

³² BMBWF, Strategie der Bundesregierung für Künstliche Intelligenz, 50.

³³ APA, Forschungsrat fordert Überarbeitung der KI-Strategie. In: derStandard.at (8.10.2024), <https://www.derstandard.at/story/3000000239824/forschungsrat-fordert-ueberarbeitung-der-ki-strategie> (10.6.2025).

3.3. KI in Schule und Unterricht

Neben den europäischen und nationalen Regelwerken wurden auch internationale Leitlinien zum Umgang mit generativer KI in der Bildung publiziert. Für die vorliegende Arbeit ist die Herausgabe der UNESCO, „Guidance for generative AI in education and research“ besonders interessant, da sie generelle Empfehlungen für den Einsatz generativer KI formuliert und dabei unter anderem Kompetenzaufbau, Governance-Fragen sowie Schutzaspekte für Schüler:innen und Lehrpersonen betont. Diese Leitlinie wurde zuletzt im April 2025 aktualisiert.³⁴ Darauf aufbauend werden im Folgenden eine Systematik sowie die Chancen und Risiken bezüglich Schule und Unterricht dargestellt.

3.3.1. Systematik, Chancen und Risiken

Neben der bereits länger andauernden digitalen Transformation lässt sich in den letzten Jahren auch eine „KI-Revolution“ beobachten, die zunehmend Auswirkungen auf den Bildungsbereich hat. Die Entwicklung von Systemen, die den Lehr- und Lernprozess sinnvoll unterstützen können, wird dabei einerseits als große Chance, aber auch als Herausforderung wahrgenommen.³⁵

Der Einsatz generativer KI im schulischen Kontext lässt sich entlang zweier Kategorien einordnen. Einerseits lässt sich innerhalb der Nutzer:innengruppe zwischen Schüler:innen und Lehrpersonen unterteilen, andererseits hinsichtlich der didaktischen Funktion im Lehr-Lern-Prozess. Durch diese Ordnung können sowohl die Potenziale als auch die damit verbundenen Risiken sinnvoll dargestellt und an konkrete Nutzungssituationen gebunden werden. Grundlegend lässt sich dabei sagen, dass die UNESCO-Leitlinie betont, dass generative KI den Lernprozess zwar unterstützen, aber nicht das menschliche Denken verdrängen sollten.³⁶

Auf der Schüler:innenseite werden von der UNESCO-Leitlinie einige Beispiele gegeben, inwiefern man generative KI sinnvoll nutzen kann. So kann ChatGPT beispielsweise zur Klärung von Begriffen, zur Strukturierung eines Themas oder zur sprachlichen Unterstützung herangezogen werden. Dabei wird betont, dass insbesondere weniger erfahrene Lernende dazu neigen können, den KI-Output ohne Prüfung zu übernehmen, auch wenn dieser oberflächlich oder fehlerhaft ist. Dementsprechend wird der Aufbau von KI- bzw. Medienkompetenzen als zentrale Voraussetzung genannt. Weiters wird auf die Nutzung von ChatGPT im Bereich Produktion und Ausdruck eingegangen. Hierzu zählen beispielsweise Gliederungsvorschläge,

³⁴ UNESCO, Guidance for Generative AI in Education and Research (Paris 2023), doi.org/10.54675/EWZM9535.

³⁵ Spyros Makridakis, The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. In: *Futures* 90 (2017) 46–60, doi.org/10.1016/j.futures.2017.03.006.

³⁶ UNESCO, Guidance for Generative AI, 25.

Formulierungs- und Schreibunterstützung oder generelle Entwürfe für Texte. Didaktisch entscheidend ist dabei, wie bereits oben erwähnt, die Grenze zwischen basaler Unterstützung und der starken Delegation von kognitiven Prozessen. Zusätzlich kann ChatGPT bei der Reflexion und Qualitätssicherung genutzt werden. Lernende können ChatGPT beispielsweise einsetzen, um Gegenpositionen zu generieren, ihre Argumente zu prüfen oder Feedback zu Entwürfen einzuholen. Die UNESCO nennt hier dialogische und projektbasierte Lernformen als mögliche Einsatzgebiete. Zuletzt ist der Bereich der Leistungsnachweise und der akademischen Integrität anzusprechen. Generative KI senkt die Hürde für Plagiate stark. Es werden zwar Detektionstools diskutiert, deren Wirksamkeit gilt jedoch als nicht hinreichend belegt. Als langfristige Konsequenz wird dementsprechend eine Weiterentwicklung von Aufgaben- und Prüfungsformaten empfohlen.³⁷

Für Lehrpersonen liegen die zentralen Potenziale von generativer KI vor allem in Unterrichtsvorbereitung und Materialerstellung. Hierbei nennt die UNESCO speziell den Nutzen bezüglich der Ideengenerierung sowie der Differenzierung und sprachlicher Vereinfachung von Aufgabenstellungen. Zusätzlich wird hervorgehoben, dass professionelle Kompetenzentwicklung nötig ist, da die Qualität der Prompts sowie die kritische Prüfung des KI-Outputs eine große Rolle dabei spielen. In der Unterrichtsdurchführung selbst kann KI als sinnvolle Lernbegleitung eingesetzt werden. Beispielsweise können vorbereitete Impulsfragen, alternative Erklärungen oder reflektierte Prompt-Strategien den Lernenden helfen, ihre Ergebnisse zu prüfen und zu begründen. Bezüglich Feedback wird von der UNESCO betont, dass KI-Vorschläge eher assistierend genutzt werden sollten und die Beurteilungsentscheidung selbst nicht an derlei Systeme delegiert werden dürfen.³⁸

Bezüglich der Risiken des Einsatzes von KI wird in der UNESCO-Leitlinie besonders vor Fehlinformationen gewarnt. KI-Tools können inhaltliche Fehler produzieren, obwohl der generierte Text gesamtheitlich korrekt und überzeugend wirkt. Dies führt eventuell zur Verbreitung von Fehlern und kann Lernprozesse verzerren. Weiters werden Datenschutz- und Bias-Risiken genannt, welche je nach Setting, in welchem die KI genutzt wird, besonders relevant werden.³⁹

Empirische Befunde aus unterschiedlichen Bildungskontexten verdeutlichen dabei, dass die Chancen und Risiken bezüglich KI weniger mit der Frequenz der Nutzung zu tun haben, sondern eher in der konkreten Art und Weise der Verwendung. So kommt eine Meta-Analyse

³⁷ UNESCO, Guidance for Generative AI, 12–33.

³⁸ UNESCO, Guidance for Generative AI, 12–28.

³⁹ UNESCO, Guidance for Generative AI, 14–18.

von Wang und Fan (2025), welche 51 Studien von 2022 bis 2025 auswertet, zu einem deutlichen Gesamteffekt von ChatGPT auf Lernleistung sowie zu mittleren Effekten auf Lernwahrnehmung und „Higher-Order-Thinking“. Diese Ergebnisse sind jedoch kontextabhängig. Die einzelnen Effekte werden stark durch Kurs- und Fachtyp, das angewandte Lernmodell, die Dauer der Intervention und die Art und Weise, wie ChatGPT genutzt wird, moderiert.⁴⁰

Auch Deng et al. (2024) berichten in ihrer Meta-Analyse über positive Effekte von ChatGPT-Interventionen auf akademische Leistung, die Motivation sowie „Higher-Order-Thinking“, während gleichzeitig eine Reduktion der mentalen Anstrengung festgestellt wird. Die Autor:innen betonen dabei, dass die untersuchten Settings im Hochschulbereich verortet sind. Dadurch ergibt sich für den schulischen Transfer die Notwendigkeit, die gemessenen Wirkungen im Kontext zu betrachten. Didaktisch besonders relevant ist der Hinweis der Studie, Leistungsüberprüfungen in KI-gestützten Settings weniger an leicht delegierbaren Aufgaben auszurichten, sondern beispielsweise Projekte einzusetzen, welche die Originalität, Kreativität und eigenen Kompetenzen der Schüler:innen sichtbarer machen.⁴¹

Auch bezüglich der in der UNESCO-Leitlinie erwähnten Plagiatsprobleme gibt es empirische Untersuchungen. Lee et al. (2024) analysieren in ihrer Arbeit beispielsweise anonyme Befragungsdaten aus drei High Schools und vergleichen Selbstberichte bezüglich Schummelns vor und nach der öffentlichen Veröffentlichung von ChatGPT. Insgesamt bleiben die berichteten Schummel-Raten dabei halbwegs stabil, jedoch zeigen sich je nach Art dabei Unterschiede, beispielsweise zwischen der Weitergabe von Informationen an Klassenkolleg:innen und dem unerlaubten Einsatz von KI. Dabei wird sichtbar, dass Schüler:innen die Zulässigkeit verschiedener KI-Nutzungen unterschiedlich bewerten. Während die Produktion einer vollständigen Arbeit oder Abgabe durch Chatbots überwiegend als nicht akzeptabel eingeschätzt wird, besteht mehr Zustimmung für die unterstützende Nutzung, etwa um einen besseren Einstieg in Aufgaben zu haben oder um neue Konzepte erklärt zu bekommen. Für den

⁴⁰ Jin Wang, Wenxiang Fan, The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. In: Humanities and Social Sciences Communications 12 (2025) 621, doi.org/10.1057/s41599-025-04787-y.

⁴¹ Ruiqi Deng, Maoli Jiang, Xinlu Yu, Yuyan Lu, Shasha Liu, Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. In: Computers & Education 227 (2025) 105224, doi.org/10.1016/j.compedu.2024.105224.

Schulkontext unterstreicht dies den Bedarf an klaren, gemeinsamen Regeln, die eine unterstützende Nutzung von völliger Delegierung der kognitiven Prozesse abgrenzen.⁴²

3.3.2. KI im österreichischen Bildungssystem

Für den österreichischen Stand liefert eine Umfrage des Österreichischen Bundesverlags ÖBV aus dem Jahr 2023 erste Einblicke. In der Umfrage (N = 334) geben 70,7 % der befragten Lehrpersonen an, dass der Umgang mit KI in der Schule erlernt werden sollte. 44,3 % nutzen im Schulalltag bereits KI-Tools, wobei ChatGPT und die Übersetzungssoftware „DeepL“ zu den am häufigsten eingesetzten zählen. Außerdem schätzen 59,9 % der Lehrpersonen den Einsatz von KI grundsätzlich als hilfreich für den Lernprozess von Schüler:innen ein. Der Mehrwert dieser Erhebung liegt zudem darin, dass sie konkrete Einsatzweisen bezüglich KI sichtbar macht. Unter jenen Lehrpersonen, die KI nutzen, wird sie von 57,1 % als Ideengeber für die Unterrichtsvorbereitung sowie von 53,7 % zur Erstellung von Arbeitsblättern verwendet. Weiters wird von 49 % der direkte Einsatz im Unterricht mit Schüler:innen genannt. Neben diesem Statusbericht zur Anwendung von KI bildet die Umfrage einen ausgeprägten Qualifizierungsbedarf ab, da sich 72,8 % der befragten Lehrpersonen mehr Angebote bezüglich der Weiterbildung beim Umgang mit KI wünschen.⁴³ Dies verdeutlicht, in welchem breiten Ausmaß die KI bereits im österreichischen Schulsystem angekommen ist und wie wichtig es ist, auf diese Entwicklungen auf allen Ebenen zu reagieren.

Neben der nationalen KI-Strategie AIMAT 2030 hat das BMBWF noch weitere Handreichungen vorgelegt, welche sich explizit mit der Digitalisierung sowie mit Künstlicher Intelligenz im Bildungsbereich befassen. Das Maßnahmenpaket „Künstliche Intelligenz – Chance für Österreichs Schulen“ stellt Ansätze vor, mit denen der Einsatz von KI im Schulwesen erprobt und begleitet werden soll. Dazu gehören etwa Pilotschulen, an denen KI-Lernsoftware getestet wird, Angebote zum Erlernen von „Prompt Engineering“, also der gezielten Formulierung von Eingaben für Chatbots wie ChatGPT, sowie der Ausbau digitaler Unterrichtsmaterialien und spezifischer Fortbildungsangebote für Lehrpersonen. Zudem wird festgehalten, dass KI bei der Erstellung schriftlicher Arbeiten grundsätzlich eingesetzt werden darf, der Umfang der Nutzung jedoch transparent gemacht werden soll.⁴⁴ Diese Aspekte der Leistungsbeurteilung und Transparenz werden in einer eigenen Broschüre weiter ausgeführt.

⁴² Victor R. Lee, Denise Pope, Sarah Miles, Rosalía C. Zárate, Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. In: Computers and Education: Artificial Intelligence 7 (2024) 100253, doi.org/10.1016/j.caeai.2024.100253.

⁴³ öbv, Wie verbreitet ist Künstliche Intelligenz an Schulen.

⁴⁴ Bundesministerium für Bildung, Wissenschaft und Forschung, Künstliche Intelligenz – Chance für Österreichs Schulen (Wien 2024), <https://www.bmbwf.gv.at/ki> (10.6.2025).

Für abschließende Arbeiten im Bereich der AHS und BMHS hat das BMBWF eine eigene Handreichung veröffentlicht. Die Herausgabe „Die Verwendung KI-basierter Tools beim Erstellen abschließender Arbeiten – Potenziale, Risiken und beurteilungsrelevante Aspekte“ verfolgt das Ziel, Betreuer:innen von abschließenden Arbeiten eine grundlegende Orientierung und eine Hilfestellung bei der Abgrenzung zwischen zulässiger Unterstützung und unzulässigem Vortäuschen von Eigenleistungen zu bieten. Ein generelles Verbot von KI-Tools wird dabei ausdrücklich als nicht sinnvoll bezeichnet, jedoch wird festgehalten, dass auch bei der Verwendung von Chatbots wie ChatGPT die Grundsätze wissenschaftlichen Arbeitens einzuhalten sind. KI-Tools dürfen dementsprechend nur unterstützend eingesetzt werden, wobei die Schüler:innen die Verantwortung für die damit erstellten Inhalte tragen und diesbezügliche Informationen eigenständig anhand geeigneter Quellen überprüfen müssen. Bezüglich der Transparenz muss im Quellenverzeichnis klar ersichtlich sein, welche KI-Systeme mit welchen Prompts im Arbeitsprozess miteinbezogen wurden. Arbeiten ohne nachvollziehbare Kennzeichnung werden als Plagiat und Täuschungsversuch gewertet.⁴⁵

Die dritte hier aufgeführte Handreichung, „Auseinandersetzung mit Künstlicher Intelligenz im Bildungssystem“, bietet einen Überblick über zentrale Begriffe wie „Künstliche Intelligenz“ und „maschinelles Lernen“ und diskutiert Chancen und Risiken der Nutzung von KI im Unterricht. Zusätzlich werden mögliche Handlungsfelder dargelegt, in denen Lehrpersonen KI einsetzen können. Die Broschüre dient Lehrpersonen als sinnvollen Einstieg in die Thematik und unterstreicht den Anspruch, KI nicht nur technisch, sondern auch pädagogisch reflektiert in das Schulsystem zu integrieren. Sie ist als Dokument angelegt, welches laufend aktualisiert und angepasst werden soll.⁴⁶

3.4. Forschungsstand zum Einsatz von ChatGPT in Bildung und Schule

Obwohl die Forschung zu LLMs und ChatGPT im Bildungsbereich noch relativ neu ist, liegen bereits mehrere Studien vor, die den Einsatz von ChatGPT in der Bildung untersuchen. Für die vorliegende Arbeit sind dabei insbesondere jene Arbeiten relevant, die die Nutzung von ChatGPT durch Schüler:innen bzw. Studierende bei der Bearbeitung fachlicher Aufgaben untersuchen. Anwendungen zur Unterrichtsplanung und Materialerstellung durch Lehrpersonen werden im Folgenden nur ergänzend erwähnt, da der Schwerpunkt der vorliegenden Arbeit auf Strategien von Schüler:innen im Umgang mit ChatGPT bei kompetenzorientierter

⁴⁵ Bundesministerium für Bildung, Wissenschaft und Forschung, Die Verwendung KI-basierter Tools beim Erstellen abschließender Arbeiten – Potenziale, Risiken und beurteilungsrelevante Aspekte. Informationen für

⁴⁶ BMBWF, Auseinandersetzung mit Künstlicher Intelligenz im Bildungssystem.

Aufgabenbearbeitung liegt. Spezifische empirische Studien zum Einsatz von ChatGPT im Geschichtsunterricht sind seltener. Diese werden im geschichtsdidaktischen Kapitel näher dargestellt.

Mehrere Studien weisen darauf hin, dass Studierende ChatGPT vor allem zur Strukturierung und Formulierung der eigenen Texte sowie zur Generierung von Anregungen zu Beginn eines Arbeitsprozesses verwenden. Wang et al. (2024) zeigen in ihrer Arbeit, dass die Teilnehmenden insbesondere Impulse und Vorlagen als hilfreich bewerten, gleichzeitig jedoch den allgemeinen Stil sowie die begrenzte inhaltliche Tiefe der Outputs der KI kritisieren.⁴⁷ Chen et al. (2024) berichten in ihrer Studie ebenfalls von positiven Effekten auf das Verständnis von Lerninhalten, wenn ChatGPT in den Lernprozess integriert wird.⁴⁸ Damit wird sichtbar, dass ChatGPT in Arbeitsprozessen sowohl als Schreib- als auch als Denkhilfe eingesetzt werden kann, der insgesamt Nutzen jedoch davon abhängt, ob Studierende die Outputs als Ausgangspunkt für eigene Arbeitsschritte nutzen oder einfach unkritisch übernehmen.

Ein didaktisch besonders interessanter Zugang findet sich in der Studie von Costa et al. (2024), die Studierende KI-generierte Antworten mit vertrauenswürdigen Quellen überprüfen lassen. Die Ergebnisse zeigen, dass sich dadurch sowohl die Motivation der Lernenden als auch deren Sensibilität für mögliche Fehlinformationen steigern lassen. ChatGPT wird in diesem Setting weniger als „Lexikon“ verstanden, sondern eher als Impulsgeber, der Reflexionsprozesse auslösen kann.⁴⁹ Gerade für kompetenzorientierte Aufgabenbearbeitung, wie in der vorliegenden Arbeit, ist dieser Befund relevant, da er zeigt, dass nicht der KI-Output selbst, sondern der Umgang mit dem KI-Output entscheidend für den Lernprozess ist.

Neben diesen Befunden werden in der Forschung wiederholt die Risiken betont, die gerade bei der Aufgabenbearbeitung und der Leistungsbeurteilung relevant werden. So weist Cooper (2023) darauf hin, dass ChatGPT häufig zu wenig transparente Belege oder Quellen anführt und daher im schulischen Bereich nur unter strenger Anleitung durch Lehrpersonen sinnvoll eingesetzt werden sollte. Zusätzlich spricht er mögliche Datenschutz- und

⁴⁷ Chuang Wang, Stephen J. Aguilar, Joseph S. Bankard, Emily Bui, Benjamin Nye, Writing with AI: What College Students Learned from Utilizing ChatGPT for a Writing Assignment. In: *Education Sciences* 14 (2024) 976, doi.org/10.3390/educsci14090976.

⁴⁸ Liang Chen, Guoliang Li, Bo Ma, Chenxi Tang, Fuhui Okubo, Atsushi Shimada, How Do Strategies for Using ChatGPT Affect Knowledge Comprehension? In: Artur M. Olney, Ioana-Andreea Chounta, Zhiyuan Liu, Olga C. Santos, Ig I. Bittencourt (Hg.), *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky* (Cham 2024) 151–162, doi.org/10.1007/978-3-031-64315-6_12.

⁴⁹ Ana R. Costa, Nuno Lima, Catarina Viegas, Ana Caldeira, Critical minds: Enhancing education with ChatGPT. In: *Cogent Education* 11 (2024) 2415286, doi.org/10.1080/2331186X.2024.2415286.

Urheberrechtsprobleme an.⁵⁰ Die Plagiatsproblematik wird dabei speziell von Khalil und Er (2023) aufgegriffen, die zeigen, dass von ChatGPT erzeugte Texte von gängigen Plagiatserkennungstools häufig nicht zuverlässig identifiziert werden können.⁵¹ Damit wird sichtbar, dass die eigentliche Herausforderung weniger in der reinen Detektion von KI-Output liegt, sondern eher in der Frage, wie zukünftige Aufgabenformate gestaltet werden müssen, damit die Eigenleistung sichtbar und beurteilbar bleibt. In diese Richtung argumentiert auch Halaweh (2023) in seiner Studie, welcher weniger einzelne Effekte untersucht, sondern Strategien für eine verantwortungsvolle Implementierung von ChatGPT formuliert. Dabei legt er den Fokus der Implementierung vor allem auf die akademische Integrität und fordert klare Rahmenbedingungen für den Einsatz in Lehre und Lernen.⁵² Zhang und Tur (2024) kommen in ihrem Review zwar zu dem Schluss, dass ChatGPT Lehrpersonen beispielsweise bei der Planung des Unterrichts und der diesbezüglichen Materialerstellung unterstützen kann, formulieren jedoch Bedenken hinsichtlich der akademischen Integrität und der Qualität von generierten Inhalten.⁵³

Ergänzend zu diesen Aspekten bearbeiten einige Studien Einsatzfelder, welche nicht unmittelbar Aufgabenbearbeitung selbst beleuchten, allerdings den schulischen Kontext einrahmen. So zeigen Jauhiainen und Guerra (2023) beispielsweise, dass von ChatGPT erstellte Lernmaterialien sinnvoll an das Leistungsniveau von Schüler:innen zwischen 8 und 14 Jahren angepasst werden können, wodurch Perspektiven für Differenzierung und Individualisierung sichtbar werden.⁵⁴ Zuletzt sei noch Uğraş et al. (2024) genannt, welche in ihrer Arbeit die Einstellungen von Lehrpersonen der Primarstufe gegenüber KI untersuchen. Sie kommen zum Ergebnis, dass diese den Einsatz von generativer KI im Unterricht grundsätzlich als sinnvoll erachten, allerdings die Problematik bezüglich Falschinformationen als zentrales Risiko wahrnehmen.⁵⁵ Diese Studien unterstreichen die spezielle Bedeutung der didaktischen Rahmung

⁵⁰ Grant *Cooper*, Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. In: *Journal of Science Education and Technology* 32 (2023) 444–452, doi.org/10.1007/s10956-023-10039-y.

⁵¹ Mohammad *Khalil*, Erol *Er*, Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection. In: Panayiotis *Zaphiris*, Andri *Ioannou* (Hg.), *Learning and Collaboration Technologies* (Cham 2023) 475–487, doi.org/10.1007/978-3-031-34411-4_32.

⁵² Mohammad *Halaweh*, ChatGPT in education: Strategies for responsible implementation. In: *Contemporary Educational Technology* 15 (2023) 421, doi.org/10.30935/cedtech/13036.

⁵³ Ping *Zhang*, Gemma *Tur*, A systematic review of ChatGPT use in K-12 education. In: *European Journal of Education* 59 (2024) e12599, doi.org/10.1111/ejed.12599.

⁵⁴ Jussi S. *Jauhiainen*, Antonio G. *Guerra*, Generative AI and ChatGPT in School Children's Education: Evidence from a School Lesson. In: *Sustainability* 15 (2023) 14025, doi.org/10.3390/su151814025.

⁵⁵ Halil *Uğraş*, Murat *Uğraş*, Stamatios *Papadakis*, Michail *Kalogiannakis*, ChatGPT-Supported Education in Primary Schools: The Potential of ChatGPT for Sustainable Practices. In: *Sustainability* 16 (2024) 9855, doi.org/10.3390/su16229855.

von KI, sie stehen jedoch in der vorliegenden Arbeit nicht im Zentrum, da hier die Strategien von Schüler:innen bei der Aufgabenbearbeitung im Vordergrund sind.

4. ChatGPT in der Geschichtsdidaktik

Das folgende Kapitel verortet ChatGPT und digitale Medien generell im geschichtsdidaktischen Rahmen kompetenzorientierten Unterrichts. Zunächst werden mit dem FUER-Modell zentrale Kompetenzdimensionen historischen Denkens erläutert. Darauf aufbauend wird Digitalisierung als Veränderung der Bedingungen für historisches Lernen diskutiert und mit geschichtsdidaktischen Systematisierungen digitaler Medien verbunden. Abschließend wird ChatGPT als spezifische Ausprägung digitaler Medien im Geschichtsunterricht anhand aktueller Studien eingeordnet.

4.1. FUER-Kompetenzen

Im aktuellen geschichtsdidaktischen Konsens zielt der Geschichtsunterricht nicht mehr darauf ab, ein „Meer“ an Informationen zu vermitteln, sondern grundlegende Kompetenzen historischen Denkens zu fördern. Dieser Ansatz ist auch im österreichischen Lehrplan für Geschichte und Politische Bildung verankert. Dort ist als Ziel formuliert, historisches und politisches Denken sowie Handeln zu vermitteln und ein reflektiertes Geschichtsbewusstsein aufzubauen. Dabei wird die Fähigkeit von Schüler:innen, zwischen Vergangenheit, Gegenwart und Zukunft zu vermitteln, als besonders wichtig hervorgehoben. Moderner Geschichtsunterricht muss lehren, sich auch in der gegenwärtigen, sehr schnell verändernden, unübersichtlichen Geschichtskultur kompetenzorientiert zu bewegen.⁵⁶

Um diesem Umstand zu begegnen, wurden Kompetenzmodelle entwickelt, welche historisches Lernen als erlernbare Fähigkeiten definiert. Schüler:innen sollen dadurch befähigt werden, beispielsweise historische Fragen zu formulieren, Quellen und Darstellungen zu dekonstruieren, historische Zusammenhänge zu rekonstruieren und daraus Orientierungen für ihr eigenes Handeln abzuleiten. Das im deutschen Sprachraum zentrale „FUER-Modell“ beschreibt die Kompetenzen des historischen Denkens daher als Frage-, Methoden-, Sach- und Orientierungskompetenz und bietet ein Strukturmodell, um diese Bereiche empirisch zu erfassen.⁵⁷

Für digitale Anwendungen, dementsprechend auch für ChatGPT als generatives KI-System, ist das FUER-Modell anschlussfähig, da digitale Werkzeuge in unterschiedlichen Phasen

⁵⁶ Andreas Körber, Kompetenzorientierung in der Domäne Geschichte (Saarbrücken 2012) 2–9; Rechtsinformationssystem des Bundes (RIS), Lehrpläne – allgemeinbildende höhere Schulen Anl. 1. In: Bundesrecht konsolidiert, tagesaktuelle Fassung, zuletzt geändert durch BGBl. II Nr. 204/2024, <https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568&Anlage=1> (21.12.2025).

⁵⁷ Waltraud Schreiber, Ein Kompetenz-Strukturmodell historischen Denkens. In: Zeitschrift für Pädagogik 54 (2008) 198–212.

historischen Denkens unterstützend eingesetzt werden können. Sie ersetzen historisches Denken nicht, sondern verändern die Bedingungen, unter denen die vier Kompetenzbereiche aktiviert werden. So beispielsweise beim Formulieren von Fragen, bei verschiedenen methodischen Arbeitsschritten der De-oder Rekonstruktion, beim Aufbau und der Strukturierung von Sachwissen sowie bei der Reflexion von Orientierungen. Weiters verschieben digitale Tools die Anforderungen an Quellen- und Darstellungskritik. Der Lehrplan nennt hierzu die Förderung kritischer Sprach- und Medienreflexion durch die Auseinandersetzung mit historischen Quellen und Darstellungen. Dementsprechend können KI-Outputs ähnlich wie Darstellungen behandelt, kritisch geprüft, kontextualisiert und in die eigene historische Argumentation eingeordnet werden.⁵⁸ Im Folgenden werden die einzelnen Kompetenzen des FUER-Modells hinsichtlich möglicher Anwendungen auf KI-generierten Output erläutert. Die Zuordnungen verstehen sich dabei als heuristische Operationalisierung des FUER-Modells für KI-gestützte Aufgabenbearbeitung.

Die „Historische Fragekompetenz“ bezieht sich auf den Umstand, dass historische Denkprozesse grundlegend mit einer diesbezüglich relevanten Frage des Individuums an die Vergangenheit beginnen, welche sich aus Interesse, Verunsicherung oder einem Orientierungsbedürfnis entwickelt. Dabei können verschiedene Kategorien von Fragen definiert werden, je nachdem, ob sich diese auf Inhalte oder Prozesse beziehen. Weiters wird grundsätzlich die Kompetenz, eigene historische Fragen zu entwickeln, von jener unterschieden, in historischen Narrationen vorliegende Fragestellungen zu erkennen. Im KI-Kontext zeigt sich die Fragekompetenz beispielsweise darin, Prompts als historische Leitfragen zu präzisieren und die Relevanz der Frage argumentieren zu können. ChatGPT kann zwar Fragevarianten oder Unterfragen vorschlagen, die Entscheidung darüber, deren Bewertung bezüglich der Sinnhaftigkeit und die dahinterliegende Argumentation liegt allerdings bei den Schüler:innen.

Die „Historische Methodenkompetenz“ bezeichnet die Fähigkeit, die im ersten Schritt aufgetauchten Fragen zu bearbeiten. Dazu gehört, vorhandene Erzählungen über Vergangenheit zu de-konstruieren und eigene historische Narrationen zu rekonstruieren. Dies kann beispielsweise durch Quellenanalysen oder Bildvergleichen verschiedenster Art geschehen. Insofern bilden die De- sowie Rekonstruktion Basisoperationen des historischen Denkens, welche sich in der konkreten Anwendung oft überschneiden. Methodenkompetente Schüler:innen können Arbeitsschritte wie Quellenkritik oder Vergleiche bewusst bei der

⁵⁸ RIS, Lehrpläne AHS Anl. 1, Anlage A (21.12.2025).

Erarbeitung eigener Darstellungen als auch bei der Analyse bereits vorhandener historischer Narrative einsetzen. Beim Arbeiten mit KI-Output wird Methodenkompetenz vor allem dort sichtbar, wo Schüler:innen KI-Aussagen oder generierte Bilder als Darstellungen behandeln und sie beispielsweise mit Quellen abgleichen. Ein zentraler didaktischer Ansatz liegt darin, dass ChatGPT methodische Schritte nur anregen darf, die eigentliche De- und Rekonstruktion aber als eigenständiger Denkprozess von den Lernenden geleistet werden muss.

Die „Historische Orientierungskompetenz“ bezieht sich auf die Fähigkeit, eigene Erfahrungen in einen größeren Sinn- bzw. Zeitrahmen einzuordnen. Historische Sinnbildungen verbinden Vergangenheit, Gegenwart und Zukunft und können damit die Bewältigung von unerwartet auftretenden Sinndiskrepanzen unterstützen. Schreiber unterscheidet dabei vier Kernkompetenzen. Die „Re-Organisationskompetenz“, welche die Reflexion des eigenen Geschichtsbewusstseins beinhaltet, eine auf Welt- und Fremdverstehen gerichtete Kompetenz, welche zur Veränderung eigener Vorstellungen von der Welt dient, die „Identitätskompetenz“, welche die Reflexion der eigenen historischen Identität betrifft, sowie historisch fundierte Handlungsdispositionen, also ein durch die Auseinandersetzung mit der Vergangenheit aufgebautes Repertoire von möglichen Handlungen. In KI-gestützten Settings wird die Orientierungskompetenz besonders dann sichtbar, wenn KI-Antworten Gegenwartsbezüge oder Handlungsimplikationen hervorrufen. Schüler:innen müssen diese Bezüge begründen und unterschiedliche Perspektiven reflektieren. ChatGPT kann alternative Deutungen oder Gegenpositionen anbieten, darf aber die eigenständigen Sinnbildungen und Verknüpfungen der Schüler:innen nicht vorwegnehmen.

Die „Historische Sachkompetenz“ schließlich bezeichnet jenen Wissensstand, welcher die Ergebnisse historischen Denkens aus den zuvor erläuterten Kompetenzen beinhaltet. Dies beinhaltet beispielsweise Kenntnisse über historische Begriffe, Konzepte oder Kategorien, mit deren Hilfe Geschichte geordnet werden kann. Sachkompetente Schüler:innen verstehen fachspezifische Begriffe, können diese von Alltagssprachlicher Verwendung unterscheiden, und historische Inhalte grundlegend strukturieren. Dabei steht nicht, wie bereits eingangs erwähnt, das bloße Faktenwissen im Mittelpunkt, sondern jenes Wissen, welches an konkrete, vorausgegangene Fragestellungen anschließt und Kommunikation über Geschichte erst ermöglicht.⁵⁹ ChatGPT kann Sachkompetenz beispielsweise unterstützen, indem es Begriffe erklärt, zeitliche Abfolgen strukturiert oder Konzepte gegenüberstellt. Dennoch obliegt es den Schüler:innen die fachliche Verlässlichkeit zu prüfen, da KI auch überzeugend klingende Fehler

⁵⁹ Schreiber, Ein Kompetenz-Strukturmodell historischen Denkens, 203–206.

produzieren kann. Sachkompetenz zeigt sich daher im KI-Rahmen nicht nur im Wissen selbst, sondern in der Fähigkeit, Begriffe und Kontextwissen aktiv zur Bewertung und Einordnung von KI-Output einzusetzen.

Die vorliegende Arbeit nutzt das FUER-Modell als zentralen Rahmen zur Analyse historischen Denkens. Die entwickelten Codes des qualitativen Teils und die darauf aufbauenden, im Datensatz gefundenen Strategietypen knüpfen an die vier Kompetenzbereiche an und erfassen, in welchen Phasen der Aufgabenbearbeitung welche Kompetenzen auftreten, unabhängig davon, ob ChatGPT genutzt wurde oder nicht.

4.2. Digitale Medien im Geschichtsunterricht

Auf Basis dieser Kompetenzen, welche bereits vor einigen Jahren ausgearbeitet wurden, stellt sich nun die Frage, wie sich die Digitalisierung des letzten Jahrzehnts auf das historische Lernen im Unterricht auswirkt. Um dies zu beantworten, wird im Folgenden Bezug auf verschiedene Ansätze der Geschichtsdidaktik genommen, in welchen die Digitalisierung generell weniger als bloßer Austausch von alten Geräten gegen neue verstanden wird, sondern als Veränderung der Bedingungen für historisches Lernen. Der Fokus liegt nicht primär auf einzelnen Tools, sondern darauf, wie die moderne digitale Umgebung Wahrnehmung, Kommunikation oder Produktion von Geschichte verändert. Historisches Lernen bleibt zwar grundsätzlich medial vermittelt, digitale Medien verschieben jedoch Reichweite, Geschwindigkeit, Darstellungsformen und Partizipationsmöglichkeiten in der Geschichtskultur. Daraus ergeben sich neue Herausforderungen an die Medien- und Urteilsfähigkeit von Schüler:innen, besonders im Hinblick auf Quellen- und Darstellungskritik.⁶⁰

Für Baumgärtner beispielsweise sind digitale Medien nicht einfach nur eine alternative Präsentationsform von Informationen, sondern ein völlig neuer Weg der Vermittlung von Inhalten, welche an bestimmte technische Voraussetzungen gekoppelt sind. Durch den Umstieg auf ein digitales Interface eröffnen sich sowohl den Schüler:innen als auch den Lehrpersonen neue Möglichkeiten, den Unterricht zu gestalten. Die einfachsten Beispiele hierfür wären die Multimodalität und die Multikodalität, also das Einwirken mittels unterschiedlicher Medien auf verschiedene Sinne. Hier sind interaktive Videos, VR und Hyperlinks zu erwähnen, welche, wenn sie korrekt in den Unterricht eingebracht werden, unterschiedlichste Interaktionen eröffnen. Zudem sind viele digitale Ressourcen für alle zugänglich und oft sogar kostenlos zu

⁶⁰ Ulrich *Baumgärtner*, *Wegweiser Geschichtsdidaktik: Historisches Lernen in der Schule* (Paderborn 2024) 249–258; Daniel *Bernsen*, Thomas *Spahn*, *Medien und historisches Lernen. Herausforderungen und Hypes im digitalen Wandel*. In: *Zeitschrift für Geschichtsdidaktik* 14 (2015) 191–203.

benutzen, wenn man von den Kosten der Hardware und den Internetgebühren sowie der Wartung der Geräte absieht. Digitale Medien treten im Unterricht dabei sowohl als Lernobjekte, beispielsweise Quellen und Darstellungen in digitaler Form, als auch als Lernprodukte, beispielsweise Präsentationen, Beiträge oder Videos, auf. Gerade diese Verschiebung hin zu stärkerer Produktion und Veröffentlichung macht digitale Lernsettings geschichtsdidaktisch relevant. Schüler:innen arbeiten nicht nur mit Materialien und neuen Geräten, sondern erzeugen zunehmend selbst Darstellungen und nehmen damit an Geschichtskultur teil.⁶¹

Im Speziellen muss hier zwischen einfacher digitaler Aufbereitung und „echten“ digitalen Angeboten unterschieden werden. Die Qualität des Lernprozesses wird beispielsweise nicht erhöht, nur weil die Lehrperson die Bildquelle, welche die Schüler:innen zu analysieren haben, mittels Beamer projiziert, anstatt Overhead-Projektor. Wenn nun allerdings Angebote von tatsächlichen digitalen Interaktionen hinzukommen, beispielsweise mittels verlinkter Erklärungen von Fachbegriffen, QR-Codes mit Erklärvideos oder 3D-Strukturen, welche am mobilen Endgerät dann noch detailreicher betrachtet werden können, wird tatsächlich eine neue Art des Lernens induziert. Solche digitalen Angebote existieren bereits, müssen allerdings in vielen Fällen erst didaktisch aufbereitet werden, da sie nicht in erster Linie für schulische Zwecke und für den Rahmen einer Unterrichtsstunde mit Eröffnung, Erarbeitung und Sicherung hergestellt wurden. Dementsprechend ersetzen Erklärvideos oder 3D-Rundgänge durch historische Stätten keinen Unterricht, aber sie können ihn in den Händen einer kompetenten Lehrperson sinnvoll ergänzen. Entscheidend ist daher weniger die technische Oberfläche, sondern eher die didaktische Rahmung. Erst wenn digitale Angebote in sinnvolle Aufgabenformate eingebettet werden, können sie historische Denkprozesse tatsächlich unterstützen.⁶²

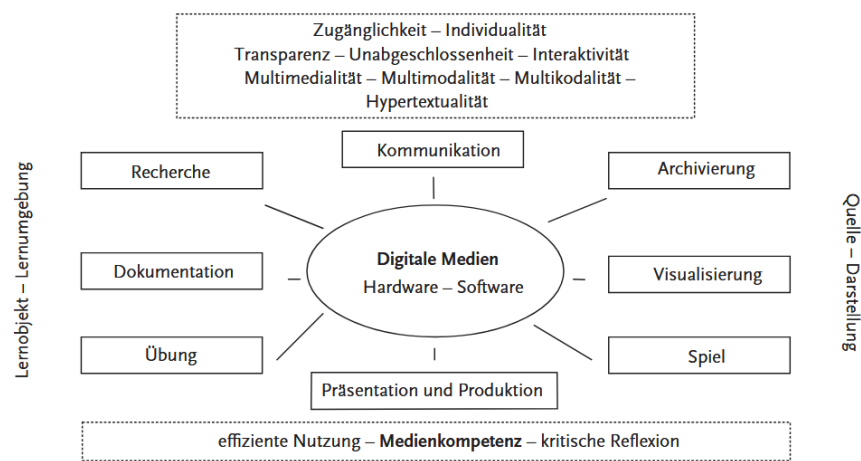


Abbildung 2: Digitale Medien im Geschichtsunterricht. Quelle: Baumgärtner, Wegweiser Geschichtsdidaktik, 256.

⁶¹ Baumgärtner, Wegweiser Geschichtsdidaktik, 249f.

⁶² Baumgärtner, Wegweiser Geschichtsdidaktik, 252f.

In der geschichtsdidaktischen Diskussion wird zur Systematisierung historischen Lernens unter Bedingungen der Digitalisierung häufig zwischen verschiedenen Modi des Lernens mit digitalen Medien unterschieden. Bernsen und Spahn differenzieren dabei „Lernen mit digitalen Medien“, „Lernen an digitalen Medien“, „Lernen über digitale Medien“ sowie „Lernen im digitalen Medium“. Baumgärtner greift diese Unterscheidung in seinem „Wegweiser Geschichtsdidaktik“ auf und visualisiert sie, wie in Abbildung 2 gezeigt. „Lernen mit digitalen Medien“ bezeichnet dabei den Einsatz digitaler Werkzeuge zur Unterstützung fachlicher Operationen, zum Beispiel bei Recherche, Strukturierung oder der Präsentation historischer Inhalte. „Lernen an digitalen Medien“ rückt dagegen die Lernangebote selbst in den Mittelpunkt, beispielsweise wenn Schüler:innen digitale Quellen- und Darstellungsformen erforschen und deren Besonderheiten wahrnehmen. „Lernen über digitale Medien“ meint die reflexive Auseinandersetzung mit Funktionsweisen, Chancen und Grenzen von digitalen Medien, beispielsweise in Hinblick auf verschiedene Perspektiven oder Interessen hinter digitalen Darstellungen. Von „Lernen im digitalen Medium“ kann schließlich gesprochen werden, wenn historische Kommunikation selbst innerhalb digitaler Räume stattfindet und Lernende dadurch unmittelbar an digitaler Geschichtskultur teilnehmen. Diese Aufteilung macht deutlich, dass digitale Medien im Geschichtsunterricht nicht nur als Hilfsmittel zur Informationsbeschaffung zu betrachten sind, sondern unterschiedliche didaktische Funktionen einnehmen können, die jeweils spezifische Anforderungen an Quellen und Darstellungen sowie an historisches Arbeiten stellen. Auch wenn die einzelnen Modi analytisch getrennt werden können, sind sie jedoch in der Praxis oft miteinander vermischt. Über diese Einteilung hinaus stellt sich jedoch die Frage, welche speziellen Kompetenzen Lernende benötigen, um in digitalen Informations- und Kommunikationsräumen historisch urteilen zu können.⁶³

Jan Hodel greift diese Frage mit dem Konzept der „Historischen Online-Kompetenz“ auf, welche historisches Lernen unter Bedingungen des Einsatzes digitaler Informationstechnologien systematisiert. Im Zentrum steht dabei nicht die alleinige Bedienung digitaler Medien, sondern die Fähigkeit, in digitalen Umgebungen historisch sinnvoll zu „lesen“, zu „schreiben“ und zu „reden“, also zu kommunizieren. Historisches „Lesen“ umfasst dabei die gezielte Informationsbeschaffung sowie die kritische Prüfung digitaler Quellen und Darstellungen, beispielsweise in Hinblick auf Perspektiven, Kontext und Verlässlichkeit. Historisches „Schreiben“ bezieht sich auf die eigenständige Erarbeitung historischer Darstellungen auf Basis ausgewählter Informationen. Historisches „Reden“ bzw.

⁶³ Baumgärtner, Wegweiser Geschichtsdidaktik, 253–256; Bernsen, Spahn, Medien und historisches Lernen, 195–200.

Kommunizieren beschreibt schließlich den Austausch über historische Deutungen, das argumentative Begründen von Urteilen sowie die reflexive Auseinandersetzung mit alternativen Sichtweisen innerhalb digitaler Kommunikationsräume. Hodel macht damit deutlich, dass historische Kompetenz im digitalen Raum nicht auf die Fähigkeit zum Einsatz von digitalen Medien allein reduziert werden kann, sondern ein Zusammenspiel aus reflexiver Bewertung, Darstellung und Urteilsbildung erfordert.⁶⁴

Im Zusammenhang mit diesen Kompetenzanforderungen diskutiert Hodel das Verhalten des „Copy & Share“, das auch in der vorliegenden Untersuchung eine zentrale Rolle spielt, da Inhalte der Konversationen mit ChatGPT während der Untersuchung häufig ohne weitere Prüfung übernommen wurden. Das Kopieren und Einfügen erscheint dabei als problematisches Verhalten, ist für Schüler:innen aber zugleich eine grundlegende Arbeitstechnik, um aus der Fülle digital verfügbarer Informationen Material für verschiedene Aufgaben wie Aufsätze oder Vorträge zusammenzustellen. Dieser Arbeitsansatz war bereits vor der digitalen Revolution vorhanden. Neu ist weniger das Sammeln von Informationen selbst, sondern die vergleichsweise einfache Übernahme von Textteilen aus sehr vielen verschiedenen digitalen Vorlagen, wodurch das Umschreiben für Lernende oft nicht sinnvoll erscheint, jedoch ein reflektierter Umgang mit den übernommenen Passagen notwendig wird. Hodel unterscheidet dabei verschiedene Formen des „Kompilierens“, was zu mosaikartigen Texten führen kann, und des „Vernetzens“, bei dem Informationen stärker miteinander verbunden werden. Diese Vorgehensweisen stehen in engem Zusammenhang mit historischen Sinnbildungsprozessen, weil Lernende bereits beim Auswählen von verschiedenen Texten Entscheidungen darüber treffen, was als relevant einzustufen ist und wie eine historische Darstellung strukturiert sein soll. Selbst beim bloßen Übernehmen von Passagen werden damit zumindest minimale narrative Leistungen sichtbar, da Auswahl, Anordnung und eventuelle Bearbeitung eine Grundstruktur voraussetzen. Aus geschichtsdidaktischer Perspektive verweist das von Schüler:innen gezeigte Kopieren und Einfügen somit darauf, dass die Recherche und anschließende Auswahl als Teil des historischen Erkenntnisprozesses verstanden werden können.⁶⁵ Auch wenn Hodel zum Zeitpunkt der Veröffentlichung von „Copy & Share“ noch nicht auf generative KI-Systeme Bezug nimmt, ist die grundlegende Argumentation dennoch in gewisser Weise übertragbar, weil ChatGPT in der Untersuchung ähnlich genutzt wurde.

⁶⁴ Jan Hodel, Historische Online-Kompetenz. In: hist.net (Archiv), Forschung & Praxis, o. J., <http://archiv.hist.net/forschung-praxis/historische-online-kompetenz/index.html> (21.12.2025).

⁶⁵ Jan Hodel, Geschichtslemen mit Copy & Share. In: hist.net – Plattform für Geschichte und digitale Medien, (11.03.2009), <https://www.hist.net/archives/2260> (21.12.2025).

Antworten der KI wurden von den Schüler:innen teilweise umformuliert, selektiv übernommen oder ohne weitere Bearbeitung in die Aufgabenlösung eingefügt.

Ein besonders naheliegendes Beispiel für veränderte Bedingungen historischen Lernens in digitalen Räumen ist Wikipedia. Lange Zeit galt Wikipedia im schulischen Kontext als besonders unzuverlässig. Jedoch hat sich die Qualität der Website in den letzten Jahren durch kontinuierliche Überarbeitungen der Artikel tendenziell erhöht, ist allerdings je nach Themenfeld diesbezüglich noch sehr heterogen.⁶⁶ Hodel selbst betont in „Wikipedia im Geschichtsunterricht“, dass Wikipedia für Schüler:innen oft die erste Anlaufstelle darstellt, wenn im Zuge eines Arbeitsauftrages neue Sachverhalte recherchiert werden müssen.⁶⁷ Diese Primärstellung könnte in den kommenden Jahren vielleicht durch Chatbots herausgefordert werden, ist jedoch derzeit noch schulische Realität. Gleichzeitig ist die Wikipedia aus geschichtsdidaktischer Sicht nicht nur eine Informationsquelle, sondern Teil einer digitalen Geschichtskultur mit eigenen Strukturen und Logiken. Dazu zählen einerseits problematische Aspekte wie der objektive Eindruck, den die Website vermittelt, sowie die stark fragmentierte Darbietung historischer Inhalte, welche in einzelne Ereignisse, Personen, oder Orte aufgeteilt und erst durch Verknüpfungen zusammengebracht werden müssen. Andererseits ergeben sich gerade daraus interessante didaktische Potenziale, weil die kritische Auseinandersetzung mit Einträgen von Wikipedia eine skeptische Grundhaltung fördern kann und die eigene Art der Auseinandersetzung mit Geschichte reflexiv thematisiert.⁶⁸ Hodel verweist hierbei auf die besondere Transparenz der Wikipedia durch Diskussionsseiten und Versionsgeschichte, anhand derer Kontroversen und Veränderungen nachvollzogen werden können. Für historisches Lernen eröffnet dies die Möglichkeit, digitale Darstellungen nicht nur zu konsumieren, sondern sie in ihrer Entstehung zu analysieren.⁶⁹ Gerade diese Transparenz fehlt Chatbots wie ChatGPT, die für Lernende damit eine „Black Box“ darstellen, wodurch ein Teil des von Hodel angesprochenen Potenzials in dieser Form nicht verfügbar ist. ChatGPT stellt dementsprechend eine noch schwierigere didaktische Herausforderung dar, da selbst Lehrpersonen keinen Zugriff auf die Trainingsdaten und mögliche Verzerrungen dieser haben.

Die bisher angeführten Ansätze zeigen, dass digitale Medien im Geschichtsunterricht nicht nur als einfache Werkzeuge, sondern als Teil der Geschichtskultur selbst verstanden werden müssen

⁶⁶ Mostafa *Mesgari*, Chitu *Okoli*, Mohamad *Mehdi*, Finn Årup *Nielsen*, Arto *Lanamäki*, The Sum of All Human Knowledge: A Systematic Review of Scholarly Research on the Content of Wikipedia. In: Journal of the Association for Information Science and Technology 66/2 (2015) 219–245, DOI: 10.1002/asi.23172.

⁶⁷ Jan *Hodel*, Wikipedia im Geschichtsunterricht (Frankfurt 2020), Klappentext.

⁶⁸ *Hodel*, Wikipedia, 7–10.

⁶⁹ *Hodel*, Wikipedia, 29–35.

und damit spezifische Kompetenzanforderungen an Recherche, Beurteilung und Darstellung entstehen. Mit generativen KI-Systemen wie ChatGPT treten diese Anforderungen noch verstärkt auf, da Informationen nicht nur gefunden oder ausgewählt, sondern in neuer sprachlicher Form erzeugt und dabei häufig ohne transparente Quellenbelege von der KI präsentiert werden. In der bildungswissenschaftlichen Diskussion werden die dafür nötigen Fähigkeiten unter dem Begriff der „AI Literacy“ zusammengefasst, welche neben den technischen Grundkenntnissen vor allem die Fähigkeit zur kritischen Prüfung, reflektierten Nutzung und Einbettung von KI-Output betonen. Für den Geschichtsunterricht stellt sich damit die Frage, wie solche Systeme eingesetzt werden und welche Strategien Lernende in kompetenzorientierten Aufgabenbearbeitungen tatsächlich zeigen. Diese Frage wird im folgenden Abschnitt anhand aktueller empirischer Studien zu ChatGPT im Geschichtsunterricht aufgegriffen.⁷⁰

4.3. ChatGPT im Geschichtsunterricht

Im folgenden Abschnitt werden drei exemplarische Studien herangezogen, welche unterschiedliche Ansätze bezüglich ChatGPT im Unterricht untersuchen und die vorliegende Arbeit hinsichtlich Struktur und Forschungsfrage beeinflusst haben. Anschließend wird auf mögliche Problematiken bezüglich des Einsatzes von generativer KI eingegangen.

Mierwald untersuchte mit „Chatting about the Past with Artificial Intelligence“ (2024) in einer qualitativen Fallstudie, wie Schüler:innen einer neunten Klasse im Rahmen einer Unterrichtseinheit mit einem Chatbot (ChatGPT 3.5) arbeiten, welcher eine historische Figur imitierte. In drei aufeinanderfolgenden Stunden bearbeiteten 21 Schüler:innen in Gruppen die Aufgabe, ein Interviewkonzept mit fünf Fragen zu entwickeln, den Chatbot anschließend entsprechend zu befragen, die Antworten zu dokumentieren und mit bereitgestellten Quellen zu vergleichen. Anschließend wertete Mierwald die Chatprotokolle der Gruppen aus. Auf dieser Grundlage entwickelt er ein Kategoriensystem, welches die wiederkehrenden kommunikativen Ansätze der Schüler:innen sowie den Output des Chatbots beschreibt. Die Analyse ergab, dass die Interaktionen mit der KI unter anderem das direkte Ansprechen der simulierten historischen Person, das Einholen von Meinungen zu historischen Themen, gezielte Nachfragen zu einzelnen Aspekten, aber auch unpräzise Fragen beinhielten. ChatGPT 3.5 reagierte in der Regel mit Antworten aus der Perspektive der historischen Person, darunter ausführlichen

⁷⁰ Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, Maggie Shen Qiao, Conceptualizing AI literacy: An exploratory review. In: Computers and Education: Artificial Intelligence 2 (2021) 100041, doi.org/10.1016/j.caeai.2021.100041.

Erklärungen und wertenden Stellungnahmen, fiel aber bei unklaren Prompts gelegentlich aus der Rolle oder produzierte Fehlinformationen. Mierwald zeigt auf, dass die Interaktion mit ChatGPT größtenteils als ein reines Frage-Antwort-Schema verlief, welches nur selten durch kritisches Nachfragen oder metakognitive Reflexionen seitens der Schüler:innen unterbrochen wurde. Es wird daher in der Studie betont, dass ChatGPT zwar neue Zugänge eröffnet, diese aber nur dann zu historischem Lernen führen, wenn die Aufgaben und die Begleitung explizit darauf ausgerichtet sind.⁷¹

Kindenberg vergleicht in „ChatGPT-Generated and Student-Written Historical Narratives“ (2024) historische Erzählungen, die von 49 Schüler:innen im Rahmen eines mehrwöchigen Projektes verfasst wurden, mit Texten, die ChatGPT-4 zu ähnlichen Aufgabenstellungen generiert. Die Teilnehmenden hatten teilweise Entscheidungsfreiheit bei der Auswahl des genauen Themas sowie des Formats und wurden während dieser Zeit speziell auf die Aufgabe vorbereitet. ChatGPT bekam ähnliche Aufgabenstellungen wie die Schüler:innen als Prompt und zusätzlich Feedback, was zu einer zweiten Version des KI-Textes führte. Die anschließende qualitative Analyse der Texte, bei welcher Kindenberg mit einem Kodierungssystem arbeitete, verglich unter anderem die narrative Qualität und das gezeigte historische Verständnis. Die Ergebnisse zeigten, dass die KI-Texte rein sprachlich zwar meist überlegen waren, hinsichtlich des historischen Verständnisses jedoch nur ungefähr das Niveau der Schüler:innen-Texte erreichten und hinsichtlich „emotional depth“ sogar oft hinter diesen zurückblieben. Kindenberg zeigt damit, dass auch „kreative“ Schreibaufgaben keinen verlässlichen Schutz vor einer KI-gestützten Textproduktion bieten, weil diese die typischen Muster von beispielsweise historischem Erzählen überzeugend imitieren kann. Dementsprechend plädiert er dafür, weniger zu versuchen, solche Plagiate zu verhindern, sondern ChatGPT gezielt im Unterricht für Aufgabendesign, Inhaltsvermittlung und Schreibcoaching zu nutzen.⁷² Damit rückt die Frage in den Vordergrund, wie Schüler:innen die KI in konkreten Aufgabenbearbeitungen tatsächlich einsetzen und welche Probleme dabei auftreten können. Genau diese Fragen sind Thema der vorliegenden Arbeit.

In der qualitativen Studie „Critical thinking and AI: Enhancing history teaching through ChatGPT simulations“ (2025) von Santamaría-Velasco et al., welche methodische Ähnlichkeiten zu jener von Mierwald aufweist, wurde untersucht, wie sich der Einsatz von

⁷¹ Marcel Mierwald, Chatting about the Past with Artificial Intelligence. A Case Study of Pupils' Interaction with ChatGPT while Completing a History-Learning Task. In: Journal of Educational Media, Memory, and Society 16 (2024) 143–173, doi.org/10.3167/jemms.2024.160206.

⁷² Björn Kindenberg, ChatGPT-Generated and Student-Written Historical Narratives: A Comparative Analysis. In: Education Sciences 14 (2024) 530, doi.org/10.3390/educsci14050530.

ChatGPT auf das historische Lernen und das kritische Denken von Studierenden auswirkt. 40 Personen verschiedener Studienrichtungen führten dazu Gespräche mit der KI, welche in der Rolle einer historischen Figur auftrat, und stellten Fragen zu historischen Ereignissen, Motiven und Folgen der Unabhängigkeitsbewegungen in Lateinamerika. Die Studierenden wurden dabei dazu aufgefordert, „deeper questions“ zu stellen und kritisch zu reflektieren, wobei ihnen einige Beispielfragen zur Verfügung standen. Die Transkripte der KI-Gespräche wurden anschließend mittels einer Inhaltsanalyse ausgewertet. Die Ergebnisse zeigen, dass die Studierenden im Verlauf der Einheit viele analytische und reflexive Denkprozesse zeigten und ein gutes Verständnis der behandelten historischen Zusammenhänge entwickelten. ChatGPT nahm in diesem speziellen Setting weniger die Rolle eines „Lexikons“ ein, sondern konnte interessante historisch-reflexive Auseinandersetzungen anstoßen.⁷³ Allerdings ist hier die Frage, welche Rolle die Kombination aus klar strukturierten Aufgaben und gezielter Steuerung durch die Lehrperson für diese Ergebnisse tatsächlich gespielt haben.

Die bisher dargestellten Möglichkeiten von LLMs im Geschichtsunterricht sollten jedoch nicht ohne die Grenzen derselben gesehen werden. Oliver Held knüpft dafür in seinem Artikel „Die Rolle des Geschichtsunterrichts in Zeiten von KI“ an die grundlegende Unterscheidung zwischen Narrativieren und Darstellen nach Handro (2015) an. Ersteres ergibt sich als ein Abschluss eines historischen Erkenntnisprozesses, welcher sowohl einem Diskurs als auch einer Überprüfung ausgesetzt ist, während letzteres der Beginn eines historischen Kommunikationsprozesses ist, welcher auf den Ergebnissen des davor abgelaufenen Prozesses beruht. LLMs halten sich aber nicht an diese „Regeln“. Sie erfahren keinen internen, von Gegenwartserfahrungen sowie Zukunftserwartungen bestimmten Sinnbildungsprozess, sondern aggregieren und rekombinieren bereits vorhandene Daten aufgrund statistischer Wahrscheinlichkeiten, ohne eigene Erfahrungen, Interessen oder Absichten dabei einzubringen. Dementsprechend plädiert Held dafür, die Outputs von LLMs wie ChatGPT nicht als „historische Darstellungen“ im engeren Sinne, sondern präziser als „Kompilationen“ oder „Rekombinationen“ zu bezeichnen.⁷⁴

Aus dieser Feststellung leitet Held mehrere Risiken für die Geschichtskultur ab. Er spricht dabei Probleme an, welche teilweise in der vorliegenden Arbeit auch im Kapitel bezüglich der technischen Grundlagen von ChatGPT, genauer gesagt bei Grenzen von LLMs, aufgebracht

⁷³ Juan Santamaría-Velasco, Aracelly Núñez-Naranjo, Ximena Morales-Urrutia, Critical Thinking and AI: Enhancing History Teaching through ChatGPT Simulations. In: International Journal of Innovative Research and Scientific Studies 8 (2025) 564–575, doi.org/10.53894/ijirss.v8i1.4403.

⁷⁴ Oliver Held, Die Rolle des Geschichtsunterrichts in Zeiten von KI. Zwischen Fakt und Fiktion: Der Einfluss von KI auf die Konstruktion und Rezeption von Geschichte. In: *Geschichte lernen* 219 (2024) 59.

werden. Da die grundlegende Zusammensetzung und die Gewichtung der Trainingsdaten für Chatbots im Normalfall nicht öffentlich einsehbar sind, lassen sich mögliche Verzerrungen des Outputs aus einer historischen Perspektive nur schwer rekonstruieren. Werden außerdem bestimmte politische oder mediale Perspektiven im Ausgangsdatensatz überrepräsentiert, so droht ein Bias der generierten Texte, welcher sich dank der sehr menschlichen Ausdrucksweise hinter dem Anschein von Objektivität verstecken kann. Besonders problematisch kann dies werden, wenn KI-Output in direkte Konkurrenz zu wissenschaftlich erarbeiteten historischen Darstellungen treten. Held warnt in diesem Kontext vor der „Realitätskonstruktion“ durch KI, die längerfristig das Vertrauen in die Wissenschaft und staatliche Institutionen untergraben kann. Aufgrund seiner Herausarbeitungen werden KI-generierte Texte in dieser Arbeit nicht als historische „Darstellungen“ im engeren Sinn, sondern als „Kompilationen“ bzw. „Rekombinationen“ bereits bestehender Darstellungen verstanden. Sie beruhen auf statistischer Verarbeitung vorhandener Texte und durchlaufen keinen eigenen, intentionalen Narrationsprozess. An Stellen, an denen andere Autor:innen den Begriff direkt verwenden, wird er im Sinne der jeweiligen Originalstudie verwendet (zB. bei Mierwald).⁷⁵

Aus diesen beiden Sichtweisen ergeben sich verschiedene Aufgaben für den Geschichtsunterricht. Auf der einen Seite eröffnen digitale Medien und mit ihnen die KI-Systeme völlig neue Zugänge zu historischen Inhalten und Möglichkeiten, mit ihnen umzugehen. Dies muss in einer geeigneten Weise von der Lehrperson in den Unterricht eingebettet und den Schüler:innen beigebracht werden, mit diesen neuen Anwendungen umzugehen. Auf der anderen Seite muss im Unterricht gelernt werden, die Outputs, die sie damit herstellen, kritisch zu hinterfragen und über den Entstehungsprozess zu reflektieren. Dementsprechend muss sowohl der Weg dahin als auch das am Ende erreichte Ziel selbst analysiert werden. Schüler:innen müssen Strategien erlernen, mit KI-Systemen umzugehen, mögliche Verzerrungen reflektieren und die damit einhergehenden Perspektiven erkennen.⁷⁶

Die vorgestellten Studien und Arbeiten zeigen, dass ChatGPT im Geschichtsunterricht grundsätzlich in unterschiedlichen Rollen sinnvoll eingesetzt werden kann. Dabei wird jedoch deutlich, dass die KI historisches Lernen sowohl unterstützen als auch behindern kann. In ungünstig strukturierten Settings tendieren Schüler:innen dazu, die KI vor allem als „Lexikon“ oder „Antwort-Generator“ zu nutzen, während mit den geeigneten Rahmenbedingungen auch analytische und reflexive Denkprozesse angeregt werden können. In den Studien bleibt jedoch

⁷⁵ Held, Die Rolle des Geschichtsunterrichts in Zeiten von KI, 59–61; König, ChatGPT und Co.

⁷⁶ Held, Die Rolle des Geschichtsunterrichts in Zeiten von KI, 59–62; Baumgärtner, Wegweiser Geschichtsdidaktik, 253–256.

offen, welche konkreten Strategien im Umgang mit ChatGPT auftreten, wenn dabei kompetenzorientierte Aufgaben zu bearbeiten sind. In diese Forschungslücke möchte die vorliegende Arbeit stoßen.

5. Methodik

Auf der Grundlage der in den vorherigen Kapiteln entwickelten und dargestellten theoretischen sowie geschichtsdidaktischen Überlegungen wird im Folgenden das methodische Vorgehen der Arbeit beschrieben.

5.1. Forschungsdesign und Forschungsfragen

Die vorliegende Arbeit untersucht in einer explorativen Weise auftretende Strategien von Schüler:innen im Umgang mit ChatGPT. Den Teilnehmenden wurden dafür mehrere kompetenzorientierte, geschichtsdidaktische Aufgaben vorgelegt, die mithilfe von ChatGPT Plus (Version 4.0) bearbeitet werden sollten. Während der Bearbeitung wurden sie gebeten, ihre Gedanken laufend zu verbalisieren („Think-Aloud“-Methode, „Lautes Denken“). Diese Äußerungen wurden mit OBS-Studio aufgezeichnet, transkribiert und anschließend qualitativ mit MaxQDA ausgewertet, um Strategietypen beim Umgang mit KI zu identifizieren.

Ergänzend zur qualitativen Datenerhebung wurde ein Fragebogen mit SoSci-Survey eingesetzt. Dieser erfasste neben demografischen Basisdaten mehrere KI-bezogene Skalen. Die Items wurden überwiegend aus bestehender Forschungsliteratur übernommen und an den schulischen Kontext angepasst. Die Antworten wurden auf Likert-Skalen erhoben und zu Mittelwerts- bzw. Summenskalen aggregiert. Die Skalen dienen dazu, den qualitativen Befunden quantitative Vergleichs- und Kontextinformationen gegenüberzustellen.

Damit wird ein Mixed-Methods-Ansatz mit qualitativem Schwerpunkt verfolgt, da die Kombination aus qualitativen und quantitativen Daten in der empirischen Forschung in vielen Fällen als besonders ertragreich gilt.⁷⁷ Einerseits werden die Interaktionen zwischen Schüler:innen und KI qualitativ analysiert, andererseits werden die Ergebnisse des standardisierten Fragebogens quantitativ ausgewertet. In einem finalen Schritt werden beide Datensätze verknüpft, indem die gefundenen Strategietypen mit den erhobenen Skalen in Beziehung gesetzt werden.

Das Studiendesign ist als explorative Querschnittsuntersuchung angelegt. Dabei werden alle Untersuchungsvariablen (quantitative als auch qualitative) zeitgleich erhoben, wodurch eine Momentaufnahme der Merkmale und ihrer Zusammenhänge ermöglicht wird.

⁷⁷ Nicola Döring, Jürgen Bortz, Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (Berlin 2023) 72f.

Die zentrale Forschungsfrage lautet:

„Welche Strategien lassen sich beim Umgang mit ChatGPT identifizieren, die Schüler:innen nutzen, um kompetenzorientierte, in den Anforderungsbereichen I–III verortete geschichtsdidaktische Aufgaben zu bearbeiten?“

Auf Basis eines deduktiv entwickelten und anschließend am Datenmaterial induktiv geschärften Kategoriensystems zu Strategien im Umgang mit ChatGPT wird anschließend geprüft, inwiefern die aufkommenden Typen mit quantitativen Merkmalen korrelieren. Daraus ergeben sich vier ergänzende Zusammenhangshypothesen:⁷⁸

„Inwiefern zeigen sich in der vorliegenden Stichprobe Zusammenhänge zwischen den definierten Strategietypen und der Nutzungsfrequenz bzw. Nutzungsdiversität von KI?“

„Inwiefern zeigen sich Zusammenhänge zwischen den Strategietypen und dem Wissen über KI und ChatGPT?“

„Inwiefern zeigen sich Zusammenhänge zwischen den Strategietypen und der Selbstwirksamkeit im Umgang mit KI?“

„Inwiefern zeigen sich Zusammenhänge zwischen den Strategietypen und den Einstellungen gegenüber KI/ChatGPT?“

Die Beantwortung der zentralen Forschungsfrage erfolgt auf Grundlage der qualitativen Typenbildung. Die vier ergänzenden Fragen werden mithilfe nichtparametrischer Korrelationsanalysen (Spearman-Rho) explorativ bearbeitet.⁷⁹

5.2. Stichprobe, Setting und Durchführung der Datenerhebung

Die gesammelte Stichprobe von N=14 besteht aus Schüler:innen des Wirtschaftsförderungsinstituts (WIFI) der Wirtschaftskammer Niederösterreichs, Standort St. Pölten. Da das WIFI eher zum Bereich der Erwachsenenbildung gehört und die Stichprobe dadurch sehr heterogen hätte ausfallen können, wurden Kriterien zur Teilnahme definiert. Die teilnehmenden Schüler:innen sollten ein Alter zwischen 16 und 22 Jahren aufweisen, um eine größtmögliche Nähe zur Zielgruppe der Sekundarstufe II herzustellen. Zum Erhebungszeitpunkt durfte die Matura noch nicht abgeschlossen sein, die Teilnehmenden sollten jedoch bereits zumindest die 9. Schulstufe besucht haben. Weiters sollten die Schüler:innen zum Zeitpunkt der Erhebung aktiv in einen oder mehrere

⁷⁸ Tobias Engelschalk, Martin Daumiller, Marion Reindl, Markus Dresel, Forschungsmethoden. In: Detlef Urhahne, Markus Dresel, Frank Fischer (Hg.), Psychologie für den Lehrberuf (Berlin/Heidelberg 2019) 533–554.

⁷⁹ Engelschalk, Forschungsmethoden, 546f.

Maturavorbereitungskurse eingeschrieben sein und diese besuchen, wobei die Fächer nicht erfragt wurden. Die Teilnahme an der Untersuchung war freiwillig.

Der Erhebungszeitraum lag zwischen dem 23.06.2025 und dem 27.06.2025. Vor Beginn der eigentlichen Datenerhebung wurde ein Pretest durchgeführt, um sowohl den Fragebogen als auch das Setting der „Think-Aloud“-Methode im Feld zu erproben. Auf Basis der Rückmeldungen aus diesem Pretest wurden kleinere Anpassungen an Instruktionen, Itemformulierungen und der technischen Umsetzung vorgenommen.

Die Datenerhebung wurde am WIFI St. Pölten in einem EDV-Raum durchgeführt. Zu Beginn wurden die Schüler:innen über den Ablauf der Studie informiert, auf die Freiwilligkeit der Teilnahme hingewiesen und die anonyme Auswertung der Daten erläutert. Anschließend erhielten alle Teilnehmenden einen individuellen Code, mit dem die quantitativen Daten später den entsprechenden qualitativen Daten zugeordnet werden konnten.

Im ersten Schritt bearbeitete jede Person den Fragebogen in SoSci-Survey allein am PC. Die Befragung umfasste demografische Angaben sowie die im nächsten Unterkapitel beschriebenen Skalen. Direkt im Anschluss wurden die qualitativen Daten erhoben: Die Schüler:innen arbeiteten in vorab gebildeten Paaren an einem PC und bearbeiteten nacheinander die drei Aufgabenblöcke. Auf den gewünschten Einsatz von ChatGPT wurde zu Beginn nochmals ausdrücklich hingewiesen.

Um eine reibungslose Datenerhebung zu gewährleisten, wurde den Schüler:innen die „Think-Aloud“-Methode kurz erläutert. In Anlehnung an die in der Literatur beschriebenen Empfehlungen, beispielsweise Sandmann (2014), bearbeiteten sie zunächst einfache Übungsaufgaben, unter anderem Kopfrechnen, um sich an das fortlaufende Verbalisieren ihrer Gedanken zu gewöhnen.⁸⁰ Erst danach starteten sie mit den geschichtsdidaktischen Aufgaben. Zu Beginn jeder Sitzung nannten die Schüler:innen laut ihren individuellen Code, sodass die Tonspur bzw. das Video später eindeutig den Fragebogendaten zugeordnet werden konnten.

Technisch wurden alle Sitzungen mit OBS-Studio aufgezeichnet. Dabei wurden sowohl der Bildschirm als auch die Audiospur der beiden Schüler:innen erfasst, um alle Interaktionen mit ChatGPT per Video nachverfolgen zu können. Die Aufnahmen wurden anschließend vollständig transkribiert und in MaxQDA importiert, wo die Kodierung nach dem

⁸⁰ Angela Sandmann, Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In: Dietmar Krüger, Ilka Parchmann, Horst Schecker (Hg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (Berlin/Heidelberg 2014) 179–188, doi.org/10.1007/978-3-642-37827-0_15.

beschriebenen Kategoriensystem erfolgte. Die Fragebogendaten wurden aus SoSci-Survey exportiert und für die quantitativen Analysen in SPSS aufbereitet.

5.3. Erhebungsinstrumente

Zur Bearbeitung der Forschungsfragen wurden zwei Instrumente eingesetzt. Einerseits qualitativ ausgerichtete Aufgabenblätter, die Schüler:innen-KI-Interaktionen anregen sollten, welche mittels der „Think-Aloud-Methode“ erhoben wurden. Andererseits ein standardisierter Fragebogen, welcher KI-bezogene Merkmale erfasste. Im Folgenden werden der Aufbau, die Funktion und die theoretische Einbettung dieser beiden Instrumente erläutert.

5.3.1. Qualitative Instrumente

Die Ansätze für die qualitativen Aufgaben, welche die Schüler:innen mittels ChatGPT zu lösen hatten, wurden aus verschiedener Fachliteratur entnommen und an die Ziele der Arbeit angepasst. Die Kernidee stammt dabei aus der Broschüre „Kompetenzorientierter Unterricht in Geschichte und Politischer Bildung: Diagnoseaufgaben mit Bildern“ von Ammerer und Windischbauer.⁸¹ Dort erwiesen sich speziell das Kapitel der Herrscherdarstellungen sowie die Anleitung zu den Anforderungsbereichen als geeignet für die Nutzung in der vorliegenden Arbeit.⁸² Alle Aufgaben sind kompetenzorientiert konstruiert, an die Anforderungsbereiche I–III des Geschichtsunterrichts ausgerichtet und im Sinne des FUER-Modells auf Sach-, Methoden-, Frage- und Orientierungskompetenz bezogen.⁸³

Im folgenden Abschnitt werden die einzelnen Aufgabenblöcke kurz erläutert und zu den Abänderungen bezüglich der Vorlage Stellung genommen. Dabei ist es wichtig anzumerken, dass von den Schüler:innen spezifisch gefordert wurde, ChatGPT bei der Bearbeitung einzusetzen. Im Anhang findet sich die gesamte, aufgegliederte Aufgabenstellung in jener Form, wie sie den Schüler:innen vorgelegt wurde. Anschließend wird die „Think-Aloud“-Methode näher erläutert.

⁸¹ Heinrich Ammerer, Elfriede Windischbauer (Hg.), *Kompetenzorientierter Unterricht in Geschichte und Politischer Bildung. Diagnoseaufgaben mit Bildern* (Wien 2011).

⁸² Irmgard Plattner, *Gemälde historischer Herrscher und Herrscherinnen als Medien der Kompetenzdiagnostik*. In: Heinrich Ammerer, Elfriede Windischbauer (Hg.), *Kompetenzorientierter Unterricht in Geschichte und Politischer Bildung. Diagnoseaufgaben mit Bildern* (Wien 2011) 42–46.

⁸³ Ulrich Trautwein, Christiane Bertram, Bodo von Borries, Nicola Brauch, Matthias Hirsch, Kathrin Klausmeier, Andreas Körber, Christoph Kühberger, Johannes Meyer-Hamme, Martin Merkt, Herbert Neureiter, Stephan Schwan, Waltraud Schreiber, Wolfgang Wagner, Monika Waldis, Michael Werner, Béatrice Ziegler, Andreas Zuckowski, *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts "Historical Thinking – Competencies in History" (HiTCH)* (Münster/New York 2017) 25–37; Ammerer, Windischbauer, *Kompetenzorientierter Unterricht*, 8f.

Der erste Aufgabenblock legt den Fokus auf die Interpretation des Herrscherportraits Ludwigs XIV. und basiert auf einer Aufgabe zu Herrschergemälden in der oben genannten Broschüre von Ammerer und Windischbauer im Kapitel „Herrschergemälde“ nach Plattner.⁸⁴ Das dort verwendete Portrait (Abbildung 3) von Hyacinthe Rigaud (1701) gilt in der geschichtsdidaktischen Literatur als klassisches Beispiel des quellenkundlichen Arbeitens mit historischen Darstellungen. Die Schüler:innen bearbeiteten dazu vier aufeinander aufbauende Fragen zu Herrschaftssymbolen, Entstehungskontext, einer leitenden Fragestellung zur Deutung des Bildes und der Einordnung der Darstellung als politische Propaganda. Die Aufgaben wurden so formuliert, dass sie die Anforderungsbereiche I–



Abbildung 3: Hyacinthe Rigaud, Ludwig XIV. von Frankreich (1701), Wikimedia Commons, https://commons.wikimedia.org/wiki/File:Louis_XIV_of_France.jpg (15.6.2025).

III abdecken. Dementsprechend reichen sie von der Beschreibung sichtbarer Bildelemente (Reproduktion, I) über die Einordnung in den historischen Kontext (Transfer, II) bis hin zur Bildung eines begründeten Urteils zur Selbstdarstellung und zum Vergleich mit modernen Darstellungen (Reflexion, III). Im Hinblick auf das FUER-Modell fokussiert sich der Block vor allem auf die Sach- und Methoden- sowie in Teilen auf die Orientierungskompetenz, wobei die Fragekompetenz nur begrenzt angesprochen wird, da der Arbeitsprozess im Wesentlichen durch vorgegebene Leitfragen strukturiert wird.

Der zweite Aufgabenblock knüpft thematisch an die Interpretation des Herrscherportraits an, fokussiert sich methodisch jedoch stärker auf die Interaktion zwischen Schüler:innen und KI. Ausgehend vom Portrait in Aufgabenblock 1 sollten die Lernenden mithilfe von ChatGPT ein eigenes Herrscher:innenbild generieren, die Eingabe an die KI schrittweise verfeinern und das daraus entstandene KI-Bild anschließend mit der historischen Darstellung vergleichen. In Aufgabe 2.1 entwickelten sie dafür einen möglichst ausführlichen Prompt und passten diesen iterativ an, wodurch sie in einen Prozess der Reflexion über die generierten Bildmerkmale hineingeführt wurden. Die Aufgabe ist vor allem in den Anforderungsbereichen II–III verortet und berührt Methoden- und Sachkompetenz, da Merkmale der Selbstdarstellung Ludwigs XIV. rekonstruiert und in konkrete Bildanweisungen übersetzt werden. Aufgabe 2.2 verlangte im Anschluss einen systematischen Vergleich zwischen dem KI-Bild und dem historischen Gemälde, was den Anforderungsbereich II anspricht, wobei insbesondere die de-konstruktive

⁸⁴ Plattner, Gemälde historischer Herrscher und Herrscherinnen, 42–46.

Analyse von Darstellungsmerkmalen und damit die Methodenkompetenz im Vordergrund stand. In Aufgabe 2.3 schließlich sollten die Schüler:innen beurteilen, ob ein in einem Geschichtsbuch abgedrucktes KI-generiertes Bild ohne Kennzeichnung problematisch wäre und ihre Position dazu begründen. Hier rückt die Quellenkritik in den Mittelpunkt, was dem Anforderungsbereich III entspricht, und es werden Methoden- sowie Orientierungskompetenz angesprochen, da der Umgang mit digital erzeugten Bildern und deren potenziell irreführender Charakter reflektiert wird.

Der dritte Aufgabenblock verlagert den Fokus von der historischen Herrscher:innendarstellung auf einen aktuelleren Anlass und knüpft an die Erinnerungskultur an. Die Schüler:innen analysierten dazu ein Werbeplakat (Abbildung 4) aus einem Tweet von Saul Sadka auf der Plattform X, auf welchem der Slogan „Never again. Until next year.“ zu lesen ist, was eine Debatte bezüglich Holocaust-Verharmlosung auslöste.⁸⁵ In Aufgabe 3.1 beschrieben die Lernenden zunächst das Plakat mit allen sichtbaren Elementen. Diese Beschreibung ist im Anforderungsbereich I verortet und bildet den Einstieg in



Abbildung 4: Saul Sadka [@Saul_Sadka], This sign is evil.... In: X (28.4.2025, 18:04), https://x.com/Saul_Sadka/status/1916886311795057089 (15.6.2025).

einen dekonstruktiven Analyseprozess, in welchem die Darstellung systematisch erfasst wird und bei dem vor allem die Methodenkompetenz angesprochen wird. In Aufgabe 3.2 sollten die Schüler:innen die Bedeutung des Slogans hinterfragen und unterschiedliche Perspektiven dazu herausarbeiten. Damit wird eine Transferleistung zwischen historischem Wissen über den Holocaust und der aktuellen Werbebotschaft verlangt, was dem Anforderungsbereich II entspricht. Neben Methoden- und Sachkompetenz wird hier weiters auch die Orientierungskompetenz angesprochen, da die Stellung des Plakats im Kontext der europäischen Erinnerungskultur reflektiert wird. Auch in diesem Aufgabenblock war der Einsatz von ChatGPT explizit gewünscht. Dies wurde bewusst so angelegt, weil der Chatbot das Plakat in den Testläufen nicht immer ohne weiteres Nachfragen angemessen historisch einordnen konnte. Gerade aus diesen Diskrepanzen heraus ergaben sich im Datenmaterial Strategien, wie die Schüler:innen KI-Antworten überprüfen und für ihre eigene Deutung des Plakats nutzbar machen.

⁸⁵ Andrew Silow-Carroll, Nike apologizes after causing uproar with 'Never Again. Until Next Year' marathon ad. In: The Times of Israel (29.4.2025), <https://www.timesofisrael.com/nike-apologizes-after-causing-uproar-with-never-again-until-next-year-marathon-ad/> (15.6.2025).

Die qualitativen Daten wurden mithilfe der „Think-Aloud“-Methode erhoben. „Lautes Denken“ gilt als sinnvolles Verfahren in der empirischen Lehr-Lernforschung, bei dem Personen ihre Gedanken während der gesamten Aufgabebearbeitung kontinuierlich verbalisieren, um kognitive Prozesse sichtbar zu machen.⁸⁶

Dabei ist es wichtig anzumerken, dass diese Methode zwar Einsicht in laufende kognitive Prozesse gewährt, allerdings auch ihre Grenzen hat. Die erfassten kognitiven Prozesse werden beispielsweise nur teilweise abgebildet, da sich komplexe oder sehr schnelle Gedankengänge sowie unbewusste Anteile nur unvollständig oder gar nicht versprachlichen lassen, was in der Literatur als „Non-Veridikalität“ bezeichnet wird. Zudem kann das Verbalisieren die Bearbeitung insgesamt verlangsamen und die Struktur des Vorgehens verändern, was als „Reaktivität“ bezeichnet wird.⁸⁷ Trotz der genannten Einschränkungen bietet die Methode eine gute Möglichkeit, im Unterschied zu rein ergebnisorientierten Verfahren auch Wahrnehmungen, Emotionen oder metakognitive Prozesse zugänglich zu machen.⁸⁸

Für die Geschichtsdidaktik ist „Think-Aloud“ dort relevant, wo historisches Denken rekonstruiert werden soll. Thyroff zeigt dabei anhand von Studien zu Ausstellungsbesuchen, dass sich mit prozessbegleitendem „Lauten Denken“ nachvollziehen lässt, wie Angebote historischer Bildung wahrgenommen und mit eigenen Erfahrungen verknüpft werden.⁸⁹ Übertragen auf die vorliegende Arbeit erlaubt die „Think-Aloud“-Methode, die Verknüpfung von Aufgabenanforderungen, Vorwissen und KI-Antworten im Verlauf der Bearbeitung sichtbar zu machen und dadurch Strategien des Einsatzes von ChatGPT sowie Ansätze historischen Denkens zu erfassen.

Während der Datenerhebung wurde nicht auf die klassische Struktur dieser Methode zurückgegriffen, sondern auf eine abgewandelte Variante, die „*Constructive Interaction*“, bei der zwei Personen gemeinsam an den Aufgaben arbeiten und ihre Überlegungen im Dialog laut aussprechen.⁹⁰ Diese Form der Methode soll einen besseren Gesprächsfluss mit tiefgreifenderen Interaktionen erzeugen und ist für die Schüler:innen weniger „unnatürlich“. Dies deckte sich auch mit den Erfahrungen aus dem Pretest der vorliegenden Arbeit, bei der einzelne teilnehmende Personen Probleme hatten, ihre Gedanken fortlaufend zu verbalisieren.

⁸⁶ Sandmann, Lautes Denken, 179.

⁸⁷ Julia Thyroff, Aneignen in einer historischen Ausstellung. Eine Bestandsaufnahme von Elementen historischen Denkens bei Besuchenden der Ausstellung «14/18. Die Schweiz und der Grosse Krieg» (Bern 2020), 203–209.

⁸⁸ Thyroff, Aneignen in einer historischen Ausstellung, 208–210.

⁸⁹ Thyroff, Aneignen in einer historischen Ausstellung, 210–216.

⁹⁰ Sandmann, Lautes Denken, 179–188; Benedikte S. Als, Janne J. Jensen, Mikael B. Skov, Comparison of Think-Aloud and Constructive Interaction in Usability Testing with Children. In: *Proceedings of the 2005 Conference on Interaction Design and Children* (Boulder 2005) 9–16.

5.3.2. Quantitative Instrumente

Die Daten für den quantitativen Teil der Arbeit wurden mittels eines standardisierten Online-Fragebogens über SoSci-Survey erhoben. Als Hauptquellen wurden die Studien von Köhler und Hartig (2024), Shahzad, Xu und Asif (2024) sowie Strzelecki (2023) herangezogen, welche sich mit dem Einsatz von KI in der Bildung befassen.⁹¹ Aus diesen wurden Skalen zu Wissen, Nutzung, Einstellungen, Selbstwirksamkeit, Vertrauen und Nutzungsfrequenz von ChatGPT übernommen und für die Schüler:innen aufbereitet. Dies wird im Folgenden erläutert. Der gesamte Fragebogen ist am Ende der Arbeit in seiner finalen Form, welche zur Datenerhebung benutzt wurde, aufgelistet.

Köhler und Hartig verwendeten in ihrer Studie Skalen, um Wissen über ChatGPT, Nutzungsmuster und Einstellungen von Studierenden zu erfassen. Für die vorliegende Arbeit wurden die Skalen „Knowledge“, „Use“ und „Attitude“ ins Deutsche übersetzt und sprachlich für die Verwendung mit Schüler:innen angepasst. Die Skala „Global“ wurde nicht übernommen, da ihre Funktion einer globalen Bewertung von KI über Items nach Strzelecki (2023) abgedeckt wurde.⁹²

Die „Knowledge“-Skala erfasst grundlegendes Wissen über Funktionsweisen von ChatGPT. Da die ursprüngliche Studie mit der Basisversion ChatGPT 3.5 arbeitete, die vorliegende Arbeit jedoch die Plus-Version ChatGPT 4.0 verwendete, wurde das Item „ChatGPT performs web searches“ entfernt, um Missverständnisse während der Durchführung zu vermeiden, da sich diese beiden Versionen dahingehend unterscheiden.⁹³ Die Schüler:innen bewerteten die Aussagen auf einer 5-stufigen Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“, ergänzt um die Ausweichoption „kann ich nicht beurteilen“. Im Unterschied zur Originalstudie, in der ein dichotomes Format („correct“ vs. „incorrect“) genutzt wurde, ermöglichte dieser Ansatz eine differenziertere Erfassung von Sicherheit bzw. Unsicherheit. Obwohl es sich bei Likert-Skalen technisch gesehen um ordinale Skalen handelt, wird in der vorliegenden Arbeit, wie in der sozialwissenschaftlichen Forschung üblich, bei mindestens fünf Antwortkategorien

⁹¹ Carmen Köhler, Johannes Hartig, ChatGPT in higher education: Measurement instruments to assess student knowledge, usage, and attitude. In: *Contemporary Educational Technology* 16 (2024) ep528, doi.org/10.30935/cedtech/15144; Artur Strzelecki, Students' Acceptance of ChatGPT in Higher Education: An Extended Unified Theory of Acceptance and Use of Technology. In: *Innovative Higher Education* 49 (2024) 223–245, doi.org/10.1007/s10755-023-09686-1; Muhammad Farrukh Shahzad, Shuo Xu, Muhammad Asif, Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model. In: *British Educational Research Journal* 51 (2025) 489–513, doi.org/10.1002/berj.4084.

⁹² Köhler, ChatGPT in higher education.

⁹³ moinAI (knowhere GmbH), ChatGPT: Die 14 wichtigsten Fragen + Antworten. In: moinAI Chatbot-Lexikon, <https://www.moin.ai/chatbot-lexikon/chatgpt> (22.6.2025).

von einem annähernd intervallskalierten Messniveau ausgegangen. Dies ermöglicht die für die Hypothesentests nötigen Berechnungen und gilt auch für alle folgenden Skalen.⁹⁴

Die „Use“-Skala erfasst, wie häufig ChatGPT für verschiedene Zwecke eingesetzt wurde. Die Items der Originalskala wurden ins Deutsche übertragen und inhaltlich an den schulischen Kontext angepasst. So wurde etwa das Item „I have used ChatGPT to have emails pre-written for me“ durch „Ich habe ChatGPT bereits genutzt, um Bilder analysieren zu lassen“ ersetzt, da diese Tätigkeit direkt an eine der im Anschluss bearbeiteten Aufgaben anknüpfte. Die Schüler:innen gaben auf einer 5-stufigen Likert-Skala von „nie“ bis „sehr häufig“ an, wie oft sie ChatGPT in der Vergangenheit für verschiedene Aktivitäten genutzt hatten. Im Unterschied zur Quellstudie, in der eine 4-stufige Skala genutzt wurde, die sowohl Nutzung als auch wahrgenommene Nützlichkeit abbildete, fokussiert die vorliegende Skala ausschließlich auf die tatsächliche Nutzungshäufigkeit. Diesbezügliche subjektive Empfindungen werden stattdessen über die anschließende Einstellungsskala erfasst.

Die „Attitude“-Skala erfasst die grundlegende Haltung gegenüber ChatGPT. Die acht Items wurden ins Deutsche übersetzt und sprachlich vereinfacht, ohne ihre inhaltliche Bedeutung zu verändern. So wurde beispielsweise „ChatGPT is not transparent with regard to the underlying text basis“ zu „Man weiß nicht genau, aus welchen Quellen ChatGPT seine Antworten zieht“ umformuliert. Die Bewertung erfolgte auf einer 5-stufigen Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“ ohne Ausweichoption. Damit wird der Ansatz der Quellstudie fortgeführt sowie gleichzeitig eine etwas feinere Differenzierung der Einstellungen ermöglicht. Außerdem sorgte dies für Einheitlichkeit mit den übrigen Skalen und erleichterte die anschließende statistische Analyse. Negativ formulierte Items wurden in SoSci-Survey direkt invertiert, sodass in SPSS ohne zusätzliche Umkodierung Mittelwerte berechnet werden konnten.⁹⁵

Ergänzend zu diesen Skalen wurden die Instrumente von Shahzad, Xu und Asif (2024) in die vorliegende Arbeit aufgenommen. In ihrer Studie untersuchen diese die individuellen Einstellungen von Studierenden gegenüber ChatGPT und erfassen dabei mehrere Konstrukte, darunter die Selbstwirksamkeit („self-efficacy“) und das Vertrauen („trust“). Beide stellten sich dabei als bedeutsame Prädiktoren der tatsächlichen Nutzung heraus und sind daher für das Verständnis der in der vorliegenden Arbeit untersuchten Strategien von besonderem Interesse.⁹⁶

⁹⁴ Engelschalk, Forschungsmethoden, 544f.

⁹⁵ Köhler, ChatGPT in higher education.

⁹⁶ Shahzad, Factors affecting generative artificial intelligence, 489.

Die Skala zur Selbstwirksamkeit erfasst die Wahrnehmung der eigenen Fähigkeiten im Umgang mit ChatGPT. Sie wurde zuerst ins Deutsche übersetzt und anschließend sprachlich an die Zielgruppe der Schüler:innen angepasst. Da die Originalitems sich an Studierende und Lehrende richteten, wurden einige Formulierungen vereinfacht. So wurde beispielsweise „Using ChatGPT for instructional purposes has become commonplace“ zu „ChatGPT im Unterricht zu verwenden ist für mich schon ganz normal geworden“ umformuliert. Die endgültige Skala umfasst sieben Items, die die Schüler:innen jeweils auf einer 5-stufigen Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“ ohne Ausweichoption zu bewerten haben.

Die „Trust“-Skala erfasst das Vertrauen der Schüler:innen in den Umgang von ChatGPT mit Informationen und Daten. Die ursprünglichen Items von Shahzad et al. wurden ins Deutsche übersetzt und sprachlich vereinfacht, ohne sie jedoch inhaltlich zu verändern. Auch hier beurteilten die Schüler:innen die sechs Items auf einer 5-stufigen Likert-Skala von „stimme gar nicht zu“ bis „stimme voll zu“, ebenfalls ohne Ausweichoption. Die ursprüngliche sechsteilige Skala wurde nach einer Faktorenanalyse auf vier Items reduziert. Die Details zur Skalenkürzung und Reliabilität werden im Abschnitt der Datenaufbereitung dargestellt.⁹⁷

Um die tatsächliche Nutzung von ChatGPT noch besser zu erfassen, wurde eine zusätzliche Variable erhoben. Strzelecki (2023) untersuchte in seiner Studie den Einsatz von ChatGPT an Hochschulen und erfasste dabei mehrere Konstrukte zu Nutzungsverhalten, Einstellungen und Akzeptanz von ChatGPT. Die Ergebnisse zeigen, dass unter anderem die Variablen Gewohnheit und Nutzungserwartung zentrale Prädiktoren für die Akzeptanz von ChatGPT sind, wobei Gewohnheit den stärksten Einfluss auf die Verhaltensabsicht und damit indirekt auf das tatsächliche Nutzungsverhalten hat. Dementsprechend wurde aus dem ursprünglichen Fragebogen die Variable „Use behavior“ übernommen, um die Skalen von Köhler und Hartig im Bereich der tatsächlichen Nutzung zu unterfüttern. Das Item wurde hinsichtlich Verständlichkeit und Anschlussfähigkeit an die Zielgruppe der Schüler:innen geprüft und in den Fragebogen integriert. Die Antwort erfolgte auf einer 7-stufigen Skala mit den Extrema „nie“ und „mehrmals am Tag“. Die Variable wird im weiteren Verlauf der Arbeit als ordinales Maß behandelt und gemeinsam mit der „Use“-Skala von Köhler und Hartig herangezogen, um ein differenzierteres Bild der tatsächlichen Nutzung von ChatGPT durch die Schüler:innen zu gewinnen.⁹⁸

⁹⁷ Shahzad, Factors affecting generative artificial intelligence, 489–513.

⁹⁸ Strzelecki, Students' Acceptance of ChatGPT in Higher Education, 223.

5.4. Datenaufbereitung und Skalenbildung

Die über SoSci-Survey erhobenen Daten wurden in IBM SPSS Statistics (Version 31) weiterverarbeitet. Zwei Testdatensätze, die nur der Erprobung des Fragebogens dienten, wurden anhand ihrer Fallnummer identifiziert und aus dem Datensatz ausgeschlossen. Anschließend wurden alle von SoSci-Survey automatisch erzeugten Metadaten gelöscht, da sie für die Beantwortung der Forschungsfragen keinen Informationsgewinn boten. In einem weiteren Schritt wurden die inhaltlich relevanten Variablen neu angeordnet und logisch gruppiert. Die Ausweichoptionen, die im Fragebogen angeboten wurden, wurden in SPSS als fehlende Werte definiert. Alle Items, die in SoSci-Survey bereits invertiert worden waren, konnten ohne weitere Umkodierung direkt verwendet werden.

Im Anschluss an die Datenaufbereitung wurden die Skalen und Kennwerte für die Hypothesentests gebildet. Die einzelnen Schritte wurden per Syntax umgesetzt. Im Folgenden wird beschrieben, wie die im vorigen Abschnitt dargestellten Skalen für die Analysen operationalisiert wurden.

Die „Knowledge“-Skala zum Wissen über KI/ChatGPT wurde berechnet, indem die fünf Items (KH01_01, KH01_02, KH01_03, KH01_05, KH01_06) dichotomisiert und in Anlehnung an die Originalstudie von Köhler und Hartig daraus ein Summenscore gebildet wurde. Bei inhaltlich falschen Aussagen bzw. Items wurde eine Ablehnung („stimme eher nicht zu“ bzw. „stimme gar nicht zu“) als korrekte Vorstellung kodiert (1), alle anderen Antworten, einschließlich „kann ich nicht beurteilen“, als falsche Vorstellung (0). Bei richtigen Aussagen wurde Zustimmung („stimme eher zu“ bzw. „stimme voll zu“) als korrekt kodiert (1), alle übrigen Antworten als inkorrekt (0). Der Wissensscore „KH01_Knowledge_Score“ ergibt sich als Summe der fünf dichotomisierten Items (Range 0–5), wobei höhere Werte für mehr Wissen stehen.

Die „Use“-Skala umfasst sieben Items (KH02_01, KH02_02, KH02_03, KH02_04, KH02_05, KH02_06 und KH02_08) zur bisherigen Nutzung von ChatGPT für verschiedene schulische Zwecke. Nach der Datenbereinigung wurde zunächst eine Reliabilitätsanalyse (Cronbachs Alpha = .90) sowie eine Faktorenanalyse mit Varimax-Rotation durchgeführt. Der erste Faktor erklärte rund 70 % der Varianz, der zweite nur ca. 19 %, sodass von einer eindimensionalen Struktur ausgegangen wurde. Dementsprechend wurde der Mittelwertindex „KH02_Use“ aus allen sieben Items berechnet. Höhere Werte stehen dabei für eine häufigere und vielfältigere Nutzung (1 = nie, 5 = sehr häufig).

Die „Attitude“-Skala beinhaltet acht Items (KH03_01 bis KH03_08), die die persönliche Einstellung zu ChatGPT erfassen sollten. Es wurde nach der Erhebung zunächst die interne Konsistenz geprüft (Cronbachs Alpha = .58). Die anschließende Faktorenanalyse mit Varimax-Rotation ergab zwei inhaltlich gut interpretierbare Faktoren (positive vs. kritische Einstellung), was die niedrige Gesamtreliabilität erklärte. Für diese wurden zwei Subskalen gebildet:

Positive Einstellung gegenüber ChatGPT: Mittelwert aus KH03_01, KH03_06 und KH03_08 („KH03_PositiveAttitude“, Cronbachs Alpha = .69), wobei höhere Werte für eine positivere Einstellung gegenüber ChatGPT stehen.

Negative Einstellung gegenüber ChatGPT: Mittelwert aus KH03_03, KH03_04 und KH03_05 („KH03_NegativeAttitude“, Cronbachs Alpha = .66), wobei höhere Werte für eine kritischere Haltung gegenüber ChatGPT stehen.

Die beiden übrigen Items (KH03_02 und KH03_07) wurden aufgrund ungünstiger Faktorladungen nicht in die Skalen aufgenommen und nur deskriptiv betrachtet.

Die „Self-Efficacy“-Skala erfasst mit sieben Items (SH01_01 bis SH01_07) die Wahrnehmung der eigenen Fähigkeiten beim Umgang mit ChatGPT. Die Skala zeigte eine sehr hohe interne Konsistenz (Cronbachs Alpha = .93). Eine Faktorenanalyse mit Varimax-Rotation ergab eine klar eindimensionale Struktur mit hohen Ladungen aller Items auf einen Faktor. Es wurde daher ein Mittelwertindex „SH01_SelfEfficacy“ aus allen sieben Items gebildet (1 = sehr geringe, 5 = sehr hohe Selbstwirksamkeit).

Die „Trust“-Skala umfasste ursprünglich sechs Items (SH02_01 bis SH02_06), welche das individuelle Vertrauen gegenüber ChatGPT abbilden sollten. Die erste Reliabilitätsanalyse ergab zwar ein gutes Cronbachs Alpha (.87), die Faktorenanalyse mit Varimax-Rotation zeigte jedoch zwei verschiedene Faktoren. Vier Items luden hoch auf einen Hauptfaktor, zwei Items (SH02_05 und SH02_06) bildeten einen zweiten, schwer interpretierbaren Faktor. Für eine möglichst homogene Skala wurden daher nur die vier Items SH02_01 bis SH02_04 berücksichtigt. Die erneute Analyse ergab eine noch bessere Reliabilität (Cronbachs Alpha = .90). Aus diesen vier Items wurde dementsprechend der Mittelwertindex „SH02_Trust“ berechnet, wobei höhere Werte gleich stärkeres Vertrauen in den sicheren und vertraulichen Umgang mit Daten durch ChatGPT bedeuten.

Die Variable „SZ03_UseBehavior“ basiert auf einem Einzelitem und misst die generelle Nutzungsfrequenz von ChatGPT auf einer 7-stufigen Skala (1 = nie, 7 = mehrmals am Tag). Das Item wurde ohne weitere Transformation übernommen. Es wird als ordinales Maß behandelt und nicht zu einer Skala aggregiert.

5.5. Strategiekategorien und Korrelationsanalyse

Die aus den „Think-Aloud“-Protokollen gebildeten Strategiekategorien wurden in SPSS als Dummy-Variablen (0 = Strategie nicht beobachtet, 1 = Strategie beobachtet) kodiert. Zusätzlich wurde ein Index über die Anzahl unterschiedlicher angewendeter Strategien pro Person gebildet. Auf Basis der Code-Dokumenten-Matrix und der Transkripte wurden anschließend drei Strategietypen gebildet, welche im Kapitel der qualitativen Ergebnisse näher erläutert werden. Diese Typenzugehörigkeit wurde paarbezogen vergeben.

Zur Untersuchung der Zusammenhänge zwischen den Strategien bzw. Strategietypen und den erhobenen Skalen wurden nichtparametrische Korrelationen mit dem Spearman-Rho-Korrelationskoeffizienten berechnet und die Skalenmittelwerte nach Strategietypen deskriptiv verglichen. Bortz und Schuster (2010) weisen darauf hin, dass Korrelationskoeffizienten in den Sozialwissenschaften ein zentrales Maß zur Beschreibung des Zusammenhangs zwischen Variablen darstellen und für unterschiedliche Skalenniveaus angepasst werden können. Da psychologische und pädagogische Merkmale häufig nicht intervallskaliert sind, empfehlen die Autor:innen ausdrücklich die Verwendung von Korrelationsmaßen, die unabhängig von Messeinheiten funktionieren und robust gegenüber unterschiedlichen Skalenniveaus sind.⁹⁹

In der vorliegenden Untersuchung kommen überwiegend ordinalskalierte Likert-Items zum Einsatz. Bortz und Schuster stellen in ihrer Ausführung bezüglich Korrelationen klar, dass für solche Daten Rangkorrelationskoeffizienten, insbesondere der Spearman-Rho, geeignet sind. Der Spearman-Koeffizient basiert auf Rangplätzen und setzt somit keine Normalverteilung oder Intervallskalierung voraus. Ein weiterer Grund für die Verwendung ergibt sich aus der geringen Stichprobengröße ($N = 14$). Parametrische Verfahren wie die Pearson-Korrelation verlangen normalverteilte und ausreichend große Stichproben, um stabile Schätzungen zu gewährleisten. Diese Voraussetzungen sind hier nicht erfüllt.¹⁰⁰

Die Berechnungen erfolgten in SPSS. Für jedes mögliche Variablenpaar wurde geprüft, ob höhere Werte in einer Skala tendenziell mit höheren oder niedrigeren Werten in der jeweils anderen Skala einhergehen. Ausgegeben wurden der Spearman-Rho-Korrelationskoeffizient sowie der zweiseitige Signifikanzwert. Fehlende Werte wurden als „Missing Values“ behandelt. Die Korrelationen wurden jeweils mit allen Personen berechnet, für die zu dem betreffenden Variablenpaar gültige Antworten vorlagen (paarweiser Fallausschluss).

⁹⁹ Jürgen Bortz, Christof Schuster, Statistik für Human- und Sozialwissenschaftler (Heidelberg 2010) 148.

¹⁰⁰ Bortz, Statistik, 176–179.

5.6. Theoretische Grundlagen der Kategorienbildung

Die Entwicklung des Kodiersystems der vorliegenden Arbeit basierte auf einem deduktiv-induktiven Vorgehen, das einerseits die KI-Nutzung im Lernkontext, andererseits geschichtsdidaktische Kompetenzmodelle berücksichtigte. Ziel war es, die Interaktionen der Schüler:innen mit ChatGPT bei der Bearbeitung von historischen Aufgaben systematisch zu erfassen und in Bezug zu geschichtsdidaktischen Aspekten zu setzen. Auf dieser Grundlage wurde ein dreidimensionales System entwickelt, welches KI-Strategien, die Nutzungsintensität und die FUER-Kompetenzen unterscheidet. Theoretisch stützt es sich vor allem auf die Arbeiten von Kim et al. (2025), Sawalha et al. (2024) und Mierwald (2024).¹⁰¹

Kim et al. (2025) analysieren in ihrer Arbeit Prompt-Muster von Studierenden bei KI-unterstützten Schreibaufgaben und zeigen, dass sich qualitativ unterschiedliche Formen der Interaktion dabei beschreiben lassen. Studierende mit hoher AI-Literacy verwenden eher kontextreiche und iterativ weiterentwickelte Prompts und setzen sich interaktiv mit den Antworten des Chatbots auseinander, während Studierende mit niedriger AI-Literacy eher allgemeine, einmalige Anfragen stellen und Output weitgehend passiv übernehmen.¹⁰² Diese Unterscheidung zwischen eher passiver, linearer Nutzung und strategischer, iterativer Nutzung bildet in der vorliegenden Arbeit die Grundlage für die Ausprägungen der „KI-Strategien“, die beispielsweise lineares vs. exploratives Prompting, Prompt-Repair sowie „KI als Informationsquelle“ vs. „KI als Denkpartner“ umfassen.

Ergänzend dazu untersuchen Sawalha et al. (2024) in ihrer Arbeit, wie Studierende ChatGPT zur Lösung technischer Aufgaben nutzen. Sie können dabei drei grundlegende Prompting-Strategien identifizieren, „Single Copy-Paste Prompting“, „Single Reformulated Prompting“ und „Multiple-Question Prompting“. Diese Strategien unterscheiden sich hinsichtlich der Komplexität, Dauer und des Reflexionsgrades der Mensch-KI-Interaktion und stehen in direktem Zusammenhang mit der Antwortqualität des Modells sowie der Prüfungsleistung der Studierenden.¹⁰³ Die Ergebnisse dienen in der vorliegenden Arbeit als Grundlage für die Ausprägung der „Intensitätsdimension“ im Kodiersystem, mit welcher erfasst wird, ob

¹⁰¹ Jinhee Kim, Seongryeong Yu, Sang-Soog Lee, Rita Detrick, Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. In: *Journal of Research on Technology in Education* (2025), doi.org/10.1080/15391523.2025.2456043; Ghadeer Sawalha, Imran Taj, Abdulhadi Shoufan, Analyzing student prompts and their effect on ChatGPT's performance. In: *Cogent Education* 11 (2024) 2397200, doi.org/10.1080/2331186X.2024.2397200; Mierwald, Chatting about the Past, 143–173.

¹⁰² Kim, Students' prompt patterns.

¹⁰³ Sawalha, Analyzing student prompts.

Schüler:innen eher mit einfachen, einmaligen Prompts arbeiten oder komplexe, iterative Interaktionen mit ChatGPT führen.

Die geschichtsdidaktische Einordnung der KI-Nutzung stützt sich schließlich auf die Arbeit von Mierwald (2024). Diese wurde bereits im Kapitel zur Geschichtsdidaktik analysiert, weswegen hier nur die Signifikanz zum Kodiersystem erwähnt wird. Er versteht KI-Antworten als historische Darstellungen, welche befragt und kritisch geprüft werden können, sodass man die Interaktion mit ChatGPT, wenn auch begrenzt, als Teil eines historischen Lernprozesses betrachten kann.¹⁰⁴ Diese Perspektive liefert für die vorliegende Arbeit die fachliche Begründung dafür, die beobachteten KI-Strategien systematisch mit den FUER-Kompetenzen zu verknüpfen.

Aus diesen drei theoretischen Strängen wurde ein dreidimensionales Kodiersystem abgeleitet:

1. KI-Strategien: In welcher Weise arbeiten Schüler:innen mit ChatGPT?
2. Nutzungsintensität: Wie tiefgehend und reflektiert wird mit der KI gearbeitet?
3. FUER-Kompetenzen: Welches historische Denken wird sichtbar?

Diese Kombination ermöglicht schlussendlich eine Analyse der Frage, welche Strategien Schüler:innen im Umgang mit ChatGPT entwickeln und wie diese Strategien mit der Qualität ihres historischen Denkens zusammenhängen.

5.7. Entwicklung und finales Kodiersystem

Die Entwicklung des Kodiersystems erfolgte in einem kombinierten deduktiv-induktiven Verfahren. Zunächst wurden exemplarisch mehrere Transkripte explorativ gelesen und erste offene Kodierungen vorgenommen, um einen Überblick bezüglich des Datenmaterials zu schaffen. Auf dieser Basis und in Rückgriff auf die beschriebenen Studien wurde das System schrittweise theoretisch angereichert, sodass alle relevanten Dimensionen der qualitativen Daten erfasst werden konnten.

Ausgehend von dieser ersten Sichtung wurde ein einfacher Codebaum entworfen, welcher sich bereits aus der KI-Ebene und der Fachebene, also dem FUER-Modell, zusammensetzte.

In einem nächsten Schritt wurden weitere Transkripte kodiert und die erste Version des Codebaums erprobt. Dabei zeigte sich, dass dieser noch nicht alle relevanten Aspekte erfasste und einzelne Elemente der Quellstudien nicht zum vorliegenden Datensatz passten. So wurde beispielsweise die soziale Kodierungsebene ergänzt, da sich in den Paaren oft klar erkennbare Rollen bzw. Arbeitsteilungen, beispielsweise „Tippen“ vs. „Ansagen“ herausbildeten und die

¹⁰⁴ Mierwald, Chatting about the Past.

Schüler:innen intensiv miteinander kommunizierten. Auch metakognitive Äußerungen, beispielsweise Planungsschritte, die Reflexion über Vorgehen und ähnliche metakognitive Äußerungen ließen sich mit dem ursprünglichen System nicht ausreichend differenziert abbilden. Diese Erweiterungen wurden in eigenen Unterkategorien innerhalb der KI-Strategien verortet.

Die FUER-Kompetenzen wurden deduktiv aus der geschichtsdidaktischen Literatur übernommen und in der ersten Testphase am Material erprobt. Es wurde geprüft, ob sie in den Äußerungen und Handlungen der Schüler:innen klar identifizierbar waren. Darauf basierend wurden für jede FUER-Kategorie typische Beispiele ausgewählt und dokumentiert, um die Zuordnung zu erleichtern und die Trennschärfe zwischen den Kategorien zu erhöhen.

Es zeigten sich bereits während dieser ersten Testphase deutliche Unterschiede zwischen den Paaren. Einige bearbeiteten die Aufgaben recht kurz, nur oberflächlich und insgesamt stark ergebnisorientiert, während andere sich deutlich mehr Zeit nahmen und iterativ, diskutierend sowie reflektierend mit ChatGPT arbeiteten. Dies führte zur Ausarbeitung einer ergänzenden Intensitätsdimension, welche bis dahin noch nicht vorgesehen war. In Anlehnung an Sawalha et al. (2024) wurden drei Niveaus definiert, welche nicht auf einzelne Äußerungen, sondern auf längere Interaktionsabschnitte, „Szenen“ bzw. Aufgabenbearbeitungen, angewendet wurden. Dabei beschreibt die erste Stufe eine sehr oberflächliche, überwiegend lineare Nutzung von ChatGPT, beispielsweise einmaliges Copy-Paste ohne Rückfragen. Die zweite Stufe bildet eine funktionale Nutzung mit Reflexions- und Anpassungsschritten ab. Unter die dritte Stufe fallen Interaktionen, welche durch iterative Nutzung, ausgedehnte Diskussionen, verschiedene Prompt-Variationen und eine tiefere Auseinandersetzung mit den KI-Antworten geprägt sind. Ergänzend wurde ein vierter Code vergeben, welcher Abschnitte kennzeichnet, in denen keine KI-Nutzung vorkommt, sodass sich im Datensatz vier Ausprägungen ergaben. Diese dienen nicht der Feinkodierung, sondern der Charakterisierung der Gesamtstrategie eines Paares innerhalb einer bestimmten Aufgabe und bilden damit eine Brücke zur späteren Strategietypenbildung.

Der Kodierungsprozess wurde in mehreren Durchläufen wiederholt, wobei die Anwendbarkeit der Kategorien jeweils geprüft und das System schrittweise verbessert wurde. In jeder Stufe wurden Verbesserungen vorgenommen, bis es als hinreichend trennscharf und handhabbar eingeschätzt werden konnte.

Das endgültige Kodiersystem setzt sich aus drei miteinander verknüpften Analyseebenen zusammen:

1. Prozessbezogene Kategorien zu KI-Strategien
2. Fachbezogene Kategorien zu den FUER-Kompetenzen
3. Vierstufige Intensitätsdimension der KI-Nutzung

Alle drei Ebenen wurden unabhängig voneinander kodiert. Mehrfachkodierungen wurden zugelassen, da sich prozessuale und fachliche Aspekte in den Äußerungen der Schüler:innen überlagern können.

5.8. Codebuch

Im Folgenden wird das für die Arbeit erstellte Kodiersystem genauer erläutert. Für jeden Code werden dabei Definitionen, typische Merkmale und ein Ankerbeispiel aus dem Datenmaterial angegeben.

Die Codes A1–A3 gehören zur Dimension der Prompting-Strategien und beschreiben diesbezügliche Formen der Formulierung und Überarbeitung von Befehlen an die KI.

A1, „Lineares Prompting“, beschreibt unveränderte Eingaben von Aufgabenstellungen oder Fragen an die KI, ohne weitere Anpassungen oder Überprüfung der Formulierung. Diese Form des Promptings tritt im gesamten Datensatz regelmäßig auf, ist die häufigste Form von Prompt-Strategien und zeigt sich besonders darin, dass Schüler:innen die gestellten Arbeitsaufträge unmittelbar an ChatGPT weitergeben und die erzeugten Antworten ohne Modifikation übernehmen.

Ein typisches Beispiel findet sich in folgendem Auszug:

8: „Schreib mal ChatGPT die Frage [2.3]. Kopiers rüber und schau ma was ChatGPT dazu sagt.“

[7 kopiert die Frage in die Prompt-Zeile von ChatGPT und schickt den Prompt ohne Veränderungen ab. Beide lesen die Antwort.]

- Paar 7–8, Zeile 82-84

Hier wird deutlich, dass die Aufgabenstellung in das Eingabefeld übertragen wird, ohne dass die Schüler:innen die Formulierung anpassen oder eine eigene Version von Frage 2.3 ausprobieren. Der Code A1 erfasst somit Situationen, in denen die KI primär als unmittelbares Antwortwerkzeug verwendet wird, ohne dass die Eingabe selbst als gestaltbarer Teil des Lösungsprozesses betrachtet wird.

A2, „Exploratives Prompting“, umfasst das bewusste Anpassen eines Prompts, bevor er an die KI gesendet wird. Im Datensatz geht dieser Code meist mit gemeinsamer Planung und Überlegung der Schüler:innen einher, welche Formulierungen geeignet sein könnten, um eine möglichst präzise oder passende Antwort zu erhalten.

Ein typisches Beispiel findet sich in folgendem Auszug:

[Beide sehen sich Frage 1.2 genauer an.]

4: „Politisch... politische Umstände.“

3: „Gib einfach ein... politische Umstände und Hintergründe des Portraits von Ludwig dem Vierzehnten. Oder du gibst einfach alles ein.“

[4 kopiert zunächst die gesamte Frage ins Prompt-Feld, beginnt aber Anpassungen daran zu machen, tippt und löscht einige Wörter.]

3: „Hintergründe‘ würde ich schreiben. Rahmenbedingungen hört sich scheiße an. Okay, wir geben jetzt ein: ‚Erklären Sie die historischen und politischen Hintergründe des Bildes.‘“

[4 gibt den überarbeiteten Prompt in die Prompt-Zeile von ChatGPT und schickt ihn ab. Beide lesen die Antwort.]

- Paar 3–4, Zeile 72–77

In dieser Sequenz wird Frage 1.2 nicht unmittelbar übernommen, sondern angepasst. Zwar ist im gewählten Beispiel der Unterschied zur originalen Aufgabe gering, dennoch zeigt es, dass sich die Schüler:innen mit dem Prompting selbst auseinandersetzen. A2 erfasst genau solche aktiven Überarbeitungsprozesse, die manchmal drastisch und manchmal eher klein ausfallen.

A3, „Iteratives Prompt-Repair“, beschreibt Überarbeitungsprozesse, bei denen Schüler:innen auf bereits erzeugte Antworten der KI reagieren und mithilfe nachgeschalteter Prompts Anpassungen, Verkürzungen oder Präzisierungen veranlassen. Im Unterschied zu explorativem Prompting erfolgt die Modifikation hier nach der ersten Rückmeldung der KI und zielt auf die Verbesserung oder Umformung der bestehenden Ausgabe ab. Ein großer Teil der in diese Kategorie fallenden Codes besteht aus Prompts an die KI, in denen diese dazu aufgefordert wird, die ausgegebene Antwort zu kürzen.

Ein typisches Beispiel findet sich in folgendem Auszug:

[Beide lesen die Antwort.]

6: „Krone auf dem Tisch... Szepter...“

5: „Ich würd's jetzt einfach kopieren und einfügen.“

6: „Ich würd's ein bisschen kürzen lassen.“

[5 gibt „kürze es bitte“ in die Prompt-Zeile von ChatGPT ein und schickt ab.]

- Paar 5–6, Zeile 18–22

In dieser Sequenz reagieren die Schüler:innen auf die von ChatGPT generierte Antwort und entscheiden, diese nicht unverändert zu übernehmen, sondern durch einen nachträglichen Prompt zu kürzen. A3 erfasst dementsprechend eine erweiterte Interaktion mit der KI nach deren erster Antwort.

Die Codes B1 und B2 erfassen, welche Rolle ChatGPT bei der Beantwortung der Fragen durch die Schüler:innen eingenommen hat.

B1, „KI als Informationsquelle“, erfasst Sequenzen, in denen die KI primär als Informationslieferant genutzt wird. In diesen Fällen dient ChatGPT als eine Art Lexikon oder Nachschlagewerk. Die Schüler:innen geben einen Prompt ein und übernehmen die ausgegebene Information ohne weitere Auseinandersetzung, Anpassung oder Interaktion. Die KI fungiert dabei nicht als dialogischer Partner, sondern ausschließlich als Quelle, aus der Inhalte extrahiert werden. B1 erfasst damit ausschließlich die Rolle der KI im Lern- bzw. Arbeitsprozess und wurde unabhängig davon kodiert, ob der Prompt zuvor überarbeitet wurde (A-Codes) oder wie intensiv die KI insgesamt genutzt wurde (S-Codes).

Ein typisches Beispiel findet sich in folgendem Auszug:

[7 kopiert Frage 1.1 in die Prompt-Zeile von ChatGPT.]

8: „Und mach in Klammer Kurzfassung, sonst schreibt er einen ganzen Roman.“

[7 ergänzt den Prompt entsprechend und schickt ihn ab. Beide lesen die Antwort.]

8: „Rüberkopieren.“

[7 kopiert die Antwort von ChatGPT ohne weitere Veränderungen in das Antwortfeld von Frage 1.1.]

- Paar 7–8, Zeile 16–18

Hier wird die von der KI ausgegebene Antwort unmittelbar in das Antwortfeld übertragen, ohne dass mit ChatGPT weiter in irgendeiner Form gearbeitet wird. Die Interaktion beschränkt sich auf das Abrufen und Übernehmen von Informationen, wodurch der Chatbot alleinig die Rolle einer externen Informationsquelle einnimmt.

B2, „KI als Partner“, umfasst Interaktionen, in denen die KI nicht nur als statische Informationsquelle genutzt wird, sondern als dialogischer Partner in den Bearbeitungsprozess einbezogen wird. In diesen Fällen reagieren die Schüler:innen auf Antworten der KI, stellen Rückfragen oder fordern Begründungen an. ChatGPT wird damit als aktiver Bestandteil des Arbeitsprozesses adressiert.

Ein typisches Beispiel findet sich in folgendem Auszug:

6: „Warum schreibt er hier auf einmal Ludwig, ich dachte es ist Louis.“
5: [lacht] „Louis wäre ein seltsamer Name.“
6: „Scroll rauf.“
[5 scrollt rauf, beide lesen die Bildunterschrift von Aufgabe 1.1]
5: „Louis!“
6: „Schau, ich hab's gesagt.“
5: „Er [der Testleiter] hat einen Fehler gemacht beim Dokumentieren.“
6: „Vielleicht weiß er das nicht.“
5: „Aber wo hat er [ChatGPT] das her [,Ludwig']?“
6: „Frag ihn. Wie er draufkommt, dass es Ludwig heißt.“
5: „Passt.“
[5 schreibt: „wie kommst du drauf das das der Ludwig ist“ und schickt ab.]
- Paar 5–6, Zeile 67–78

In dieser Sequenz reagieren die Schüler:innen unmittelbar auf eine Inkonsistenz in der KI-Antwort und nutzen ChatGPT aktiv zur Klärung der Frage. Die KI wird dabei nicht nur als Quelle einer Information betrachtet, sondern als Akteur, dessen Entscheidungsweg und Zuordnung hinterfragt wird. B2 bildet dementsprechend eine Art Aufstufung der KI von einer Art „Informations-Blackbox“ zu einem Gesprächspartner ab, welcher mehr in die Aufgabenlösung miteinbezogen wird.

Die Codes C1–3 befassen sich mit den metakognitiven Prozessen der Schüler:innen, welche beim Bearbeiten der Aufgaben sichtbar wurden.

C1, „Planung des Vorgehens“, umfasst Sequenzen, in denen das weitere Vorgehen durchdacht, Handlungsoptionen verglichen oder beispielsweise eine Arbeitsreihenfolge festgelegt wurde. Diese metakognitive Steuerung tritt oft früh in der Bearbeitungsphase der einzelnen Aufgaben quer über alle Paare hinweg auf.

Ein typisches Beispiel findet sich in folgendem Auszug:

[Beide lesen die Aufgabe.]
7: „Soll ich das [Frage 1.1] mal rauskopieren? [In die Prompt-Zeile von ChatGPT]“
8: „Schauen wir uns zuerst das Bild an, oder?“
7: „Ok.“
- Paar 7–8, Zeile 4–7

In dieser Sequenz diskutieren die Schüler:innen zwei mögliche Einstiege in die Bearbeitung und einigen sich anschließend auf eine Reihenfolge. Planung kann dabei auch ohne eine explizite Diskussion auftreten, wenn beispielsweise nur eine Person die weiteren Arbeitsschritte vorgibt. C1 erfasst dementsprechend generelle Planungs- und Koordinationsschritte, in denen der Arbeitsprozess bewusst strukturiert wird.

C2, „Bewertung“, umfasst Sequenzen, in denen Schüler:innen die von der KI gelieferten Antworten inhaltlich einschätzen, kommentieren oder gegeneinander abwägen. Diese Bewertungen beziehen sich sowohl auf die Plausibilität als auch auf die Passung der KI-Antwort zur gestellten Aufgabe. In vielen Fällen gehen Bewertung und Auswahl eng miteinander einher, wenn beispielsweise bestimmte Elemente übernommen, andere hingegen verworfen werden.

Ein typisches Beispiel findet sich in folgendem Auszug:

[Beide betrachten die Antwort von ChatGPT und kommentieren diese.]

2: „Gut, also statt Richterstab, Zepter. Dann Hand der Gerechtigkeit schreiben wir noch dazu. Hermelin-Mantel habe ich nicht gewusst. Königliche Würde, Reichtum und Tradition.“

1: „Die Pose ist mir auch schon suspekt vorgekommen.“

2: „Die Pose würde ich jetzt streiten, aber ist okay. Okay, wir nehmen sie.“

[2 schreibt in eigenen Worten die von ChatGPT aufgezählten Herrschaftssymbole ins Antwortfeld.]

- Paar 1–2, Zeile 13–17

In dieser Sequenz gehen die Schüler:innen die von ChatGPT genannten Symbole nacheinander durch, kommentieren diese und äußern sich zustimmend oder leicht distanziert. Zugleich treffen sie eine Auswahl, welche Aspekte in die eigene Antwort übernommen werden sollen. C2 erfasst dementsprechend Phasen der inhaltlichen Bewertung von KI-Antworten.

C3, „Strategiewechsel“, erfasst Sequenzen, in denen Schüler:innen ihre ursprüngliche Vorgehensweise unterbrechen, verwerfen oder anpassen und eine neue Strategie wählen. Strategiewechsel können dabei sowohl die eigene Bearbeitung betreffen als auch den Umgang mit ChatGPT.

Ein typisches Beispiel findet sich in folgendem Auszug:

[2 tippt in das Antwortfeld von Frage 3.1 „Zeigt einen humorvollen/motivierenden Spruch, der sich auf das Laufen od. einen Marathon bezieht.“]

2: „Nein, weißt du was, der [ChatGPT] soll mir das zusammenfassen.“

1: „Ja, lass es ChatGPT zusammenfassen.“

2: „Auf 6 Zeilen.“

1: „Ja.“

[2 tippt in die Prompt-Zeile von ChatGPT „fasse kürzer zusammen, auf 6 zeilen, ohne absätze“. 2 löscht seine eigene Antwort aus dem Antwortfeld.]

- Paar 1–2, Zeile 168–175

Hier wird für die Aufgabe 3.1 eine Antwort eingetippt, welche auf einer vorigen KI-Antwort basiert, allerdings selbst formuliert wird. Anschließend beschließen die Schüler:innen jedoch, mittels einer weiteren KI-Interaktion ChatGPT damit zu beauftragen, seine vorher gegebene Antwort zusammenzufassen. C3 erfasst dementsprechend Übergänge zwischen verschiedenen Arbeitsansätzen.

Die Codes D1 und D2 beschreiben soziale Prozesse, welche zwischen den Schüler:innen während der Aufgabenbearbeitungen auftraten.

D1, „Rollenverteilung“, erfasst Sequenzen, in denen Schüler:innen explizit Arbeitsaufgaben oder Verantwortlichkeiten untereinander aufteilen. Diese Rollenverteilung umfasst im Datensatz meistens Entscheidungen darüber, wer schreibt, wer Inhalte diktiert, wer Interaktionen mit der KI übernimmt oder wer Bewertungen ausspricht. Die Rollen entstehen situativ im Gespräch und strukturieren den Bearbeitungsprozess.

Ein typisches Beispiel findet sich in folgendem Auszug:

3: „Die Fachwörter kannst du reinschreiben, Digga.“

4: „Ach so, ich muss schreiben.“

[4 beginnt in eigenen Worten die von ChatGPT aufgezählten Herrschaftssymbole ins Antwortfeld zu schreiben.]

3: „Soll ich dir ansagen?“

4: „Ja, sag an.“

[4 schreibt weiter, während 3 ansagt.]

- Paar 3–4, Zeile 42–47

In dieser Sequenz wird die Aufgabenverteilung direkt ausgehandelt. Eine Person übernimmt das Schreiben, während die andere diktiert. Die Rollen sind damit klar definiert und werden ohne weitere Diskussion umgesetzt. D1 bildet dementsprechend Momente der Arbeitszuweisung und arbeitsteiligen Kooperation ab.

D2, „Aushandlung und Entscheidung“, umfasst Sequenzen, in denen Schüler:innen verschiedene Vorschläge oder Formulierungsoptionen gemeinsam diskutieren, gegeneinander

abwägen und schließlich zu einer konsistenten Entscheidung gelangen. In den meisten Fällen wird im vorliegenden Datensatz darüber gesprochen, was geschrieben werden soll und wie es formuliert werden soll.

Ein typisches Beispiel findet sich in folgendem Auszug:

3: „Okay, wie fassen wir das in Worte? Sonst schreib vielleicht einfach so ‚Vergleich zu moderner Darstellung‘. Weil ich glaube, das ist schwierig, das in einem entscheidenden Satz zu verpacken.“

[4 beginnt im Antwortfeld zu schreiben, 3 diktiert teilweise.]

3: „Sollen wir in einer Klammer vielleicht ein paar Beispiele dazu schreiben, dass man sich das dahinter vorstellen kann? Wie jetzt zum Beispiel Kleidung oder Farbe?“

4: „Ich würde sagen, Farbe ist einfach ganz wichtig. Weil bei so Wahlplakate ja auch mit Farbe gespielt wird, damit du weißt, wo du hinschauen musst.“

3: „Sehr gut. Schreib Farbe.“

- Paar 3–4, Zeile 167–171

In dieser Passage entwickeln die Schüler:innen gemeinsam eine geeignete Formulierung, diskutieren mögliche Beispiele und entscheiden schließlich, den Aspekt „Farbe“ explizit zu integrieren. D2 erfasst dementsprechend kooperative Abstimmungs- und Entscheidungsprozesse, in denen die inhaltliche Gestaltung des Textes gemeinschaftlich verhandelt wird.

E1, „Bedienung“, umfasst alle Hinweise auf technische Bedienhandlungen, die den Umgang mit der KI-Oberfläche betreffen. Dazu zählen unter anderem das Kopieren, Einfügen, Absenden von Prompts sowie das Übertragen von KI-Inhalten in das Antwortfeld von Aufgaben. E1 kodiert dabei ausschließlich die technische Ebene der Interaktion und wird unabhängig von inhaltlichen oder strategischen Aspekten (A–D-Codes, S-Codes) vergeben.

Ein typisches Beispiel findet sich in folgendem Auszug:

2: „Und dann kopieren wir den Wisch und hauen es rein.“

[Bild aus Aufgabe 1.1 wird kopiert und in die Prompt-Zeile von ChatGPT eingefügt. Das eingefügte Bild wird mit der Beifügung „Herrschaftssymbole“ in der Prompt-Zeile abgeschickt.]

- Paar 1–2, Zeile 8–9

In dieser Sequenz kopieren die Schüler:innen ein Bild, fügen es in das Eingabefeld von ChatGPT ein und senden die Kombination mit einer kurzen Zusatzanweisung ab.

Die Codes F1–4 umfassen die FUER-Kompetenzen, die bei der Bearbeitung der historischen Aufgaben auftraten.

F1, „Sachkompetenz“, umfasst Sequenzen, in denen Schüler:innen historisches Sachwissen aktivieren, Begriffe korrekt verwenden oder fachliche Zuordnungen vornehmen. Dazu gehört die Benennung historischer Personen, Konzepte oder Symbole sowie das Erkennen und Einordnen zentraler Kategorien des historischen Kontextes. Sachkompetenz zeigt sich im vorliegenden Datensatz meistens, wenn Schüler:innen ihr Wissen nutzen, um historische Darstellungen zu identifizieren oder einzuordnen.

Ein typisches Beispiel findet sich im folgenden Auszug:

8: *„Schauen wir uns zuerst das Bild an, oder?“*

7: *„Ok.“*

8: *„Das is Ludwig der 14. Auch Sonnenkönig genannt.“*

- Paar 7–8, Zeile 6–8

In dieser Sequenz identifiziert eine Person die dargestellte Person korrekt als Ludwig XIV. und ergänzt die historische Bezeichnung „Sonnenkönig“. Damit werden sowohl fachliche Begriffe als auch historisches Kontextwissen einbezogen.

F2, „Methodenkompetenz“, umfasst Sequenzen, in denen Schüler:innen historische Darstellungen methodisch erschließen, indem sie Merkmale identifizieren, zuordnen oder interpretieren. Dazu gehört das Beurteilen von Bildelementen, das Vergleichen von Darstellungen sowie das Heranziehen eigener Kenntnisse, um historische Symbole zu erkennen oder zu deuten. Methodenkompetenz zeigt sich im Datensatz meistens bei der Arbeit mit den Bildquellen.

Ein typisches Beispiel findet sich im folgenden Auszug:

4: *„Ich würde eigentlich das Symbol [4 deutet mit dem Mauszeiger auf die Lilie am Umhang] mal als Lilie werten. Zumindest kenne ich es von den Pfadfindern.“*

[Beide lachen.]

- Paar 3–4, Zeile 27

In dieser Sequenz wird ein Symbol im Gemälde identifiziert und mithilfe eigenen Vorwissens eingeordnet. Während der Bezug humorvoll ist, beschreibt die Äußerung dennoch einen methodischen Analysevorgang, nämlich die Identifizierung und Einordnung eines historisch relevanten Symbols.

F3, „Fragekompetenz“, umfasst Sequenzen, in denen Schüler:innen historische Fragestellungen entwickeln, formulieren oder reflektieren. Fragekompetenz zeigt sich besonders dann, wenn eigene Fragen formuliert werden, die ein historisches Problem sichtbar machen und damit einen Denkprozess in Gang setzen. Im Datenmaterial zeigt sich Fragekompetenz vor allem bei der Analyse der Bildquellen.

Ein typisches Beispiel findet sich im folgenden Auszug:

[An 2 gewandt]

1: „Was ist deine Frage?“

2: „Was ist meine Frage? Hm. Selbstdarstellung von Ludwig dem 14.“

[2 gibt ins Antwortfeld ein und liest mit: „Wie lässt sich ein luxuriöser Kleidungsstil und daraus resultierender Lebensstil mit den christlichen Werten, auf der die Legitimation der Herrschaft aufgebaut ist, vereinbaren?“]

- Paar 1–2, Zeile 55–58

In dieser Sequenz wird eine eigene historische Fragestellung entwickelt, die auf der Diskrepanz zwischen der Selbstdarstellung Ludwigs XIV. und den christlichen Werten, auf die dieser seine Legitimation stützt, abzielt. Die Frage verbindet Beobachtungen aus der Quelle mit einem historischen Problemfeld und initiiert einen analytischen Denkprozess.

F4, „Orientierungskompetenz“, umfasst Sequenzen, in denen Schüler:innen Bezüge zwischen Vergangenheit und Gegenwart herstellen, historische Darstellungen zur Reflexion über gesellschaftliche Zusammenhänge nutzen oder politische Kontexte kritisch einordnen.

Ein typisches Beispiel findet sich im folgenden Auszug:

[2 kopiert Frage 2.3 in die Prompt-Zeile von ChatGPT.]

2: „Mich würde interessieren, was in China herauskommt, wenn ich das eingebe. Wenn ich in China etwas zum Massaker am Tian’anmen-Platz eingebe, kommt zum Beispiel nichts.“

– Paar 1–2, Zeile 150–151

In Frage 2.3 geht es darum, inwiefern das Selbstportrait Ludwigs XIV. als Propaganda gewertet werden kann. In diesem Zusammenhang reflektieren die Schüler:innen über die politische Steuerung digitaler Informationen und stellen einen Gegenwartsbezug her, indem sie auf die chinesische Internetzensur verweisen. Die Äußerung stellt damit eine Brücke zwischen der historischen Thematik und moderner Informationskontrolle dar.

Die Codes S0–S3 erfassen die Intensität, mit der die KI in den Bearbeitungsprozess eingebunden wurde.

S0, „Keine KI-Nutzung“, umfasst Sequenzen, in denen Schüler:innen die KI nicht zur Bearbeitung der Aufgabe verwenden. Häufige Gründe im Datensatz waren dabei bewusster Verzicht auf Hilfe von ChatGPT, weil die Aufgabe auch ohne KI lösbar war oder weil die eigenständige Beantwortung als einfacher erschien.

Ein typisches Beispiel findet sich im folgenden Auszug:

[Lesen Aufgabe 1.4]

1: „Was ist Propaganda?“

2: „Wenn du es unter das Volk bringst.“

1: „Frag mal Chat-GPT.“

2: „Wieso, das kann ich auch so beantworten.“

1: „Okay.“

– Paar 1–2, Zeile 61–66

In dieser Sequenz wird auf ChatGPT verzichtet, da die eigene Kompetenz als hoch genug eingeschätzt wird, die Frage auch ohne Hilfe beantworten zu können. Die KI wird somit bewusst nicht in die Bearbeitung einbezogen.

S1, „Oberflächliche KI-Nutzung“, umfasst Sequenzen, in denen Schüler:innen ChatGPT zwar verwenden, dies jedoch ohne großartige inhaltliche Auseinandersetzung, ohne iterative Anpassungen und ohne gezielte Steuerung der Antwort tun. Charakteristisch ist ein mechanisches Abschicken der kopierten Aufgabe, häufig ohne die Fragestellung vollständig zu lesen. Die KI-Antwort wird anschließend unreflektiert übernommen. S1 markiert damit ein sehr niedrigschwelliges Nutzungsniveau, das sich auf das reine Kopieren und Einfügen beschränkt.

Ein typisches Beispiel findet sich im folgenden Auszug:

[Ohne die Frage gelesen zu haben]

14: „Ich kopiers wieder rein.“

[Frage 1.3 wird ohne Abänderung in die Prompt-Zeile kopiert und abgeschickt.]

[14 liest erst jetzt die Frage, dann die Antwort von ChatGPT.]

14: „Sehr stark, sind zwei gute Fragen.“

[Die Antwort wird 1:1 in das Antwortfeld 1.3 kopiert.]

- Paar 13–14, Zeile 31–36

In dieser Sequenz wird die KI genutzt, ohne zuvor die Aufgabenstellung zu prüfen oder den Prompt inhaltlich zu verändern. Die KI-Antwort wird anschließend unverändert übernommen.

S2, „Funktionale Nutzung“, umfasst Sequenzen, in denen Schüler:innen ChatGPT gezielt zur Unterstützung der Aufgabenbearbeitung einsetzen, dabei aber oft eigene Prompts schreiben oder die KI-Antworten nicht einfach in das Antwortfeld kopieren, sondern funktional in den eigenen Arbeitsprozess integrieren. Die KI dient damit als Werkzeug, nicht als vollständiger Ersatz für eigene Denkleistungen.

Ein typisches Beispiel findet sich im folgenden Auszug:

[Beide betrachten die Bildquelle aus Aufgabenblock 1.]

10: „Ich mein, wir könnten versuchen, ihn [ChatGPT] das Bild beschreiben zu lassen, das kopieren wir dann einfach da unten rein [Prompt-Zeile von ChatGPT]. Oder wir schreiben es selbst. Je nachdem, was mehr Arbeit wäre.“

9: „Dann machen wir das andere.“

10: „Okay.“

[10 kopiert die Bildquelle aus Aufgabe 1 in die Prompt-Zeile von ChatGPT, tippt „beschreibe das bild detailliert“ und schickt ab. ChatGPT gibt eine längere Beschreibung.]

9: „Wir kopieren das und ändern es einfach ein wenig um.“

10: „Ja, das ist gut.“

[10 beginnt mit Hilfe von 9 Auszüge aus der Antwort ChatGPTs in die Prompt-Zeile zu kopieren und formuliert sie um, indem er den Namen Ludwigs streicht und „ein Herrscher“ schreibt. Zusätzlich stellt er im Aufgabenbereich der KI auf „Erstelle ein Bild“ um. ChatGPT erstellt ein täuschend ähnliches Bild.]

- Paar 9–10, Zeile 51–60

In dieser Passage entscheiden die Schüler:innen, dass ChatGPT die Bildbeschreibung verfassen soll. Sie nutzen anschließend die von der KI generierte Beschreibung als Grundlage, übernehmen Teile daraus und passen den Text an ihre Zwecke an. Anschließend verwenden sie die KI, wie in Aufgabe 2 gefordert, um auf Basis der Beschreibung ein neues Bild erstellen zu lassen.

S3, „Reflexive, iterative Nutzung“, umfasst Sequenzen, in denen Schüler:innen ChatGPT wiederholt in einen zyklischen Bearbeitungsprozess einbinden, die KI-Antworten kritisch beurteilen und durch gezielte Promptveränderungen weiterentwickeln. Charakteristisch ist ein reflexives Vorgehen, bei dem die Lernenden eigene Ideen mit KI-Rückmeldungen verknüpfen, neue Varianten anfordern und auf Basis der Ergebnisse weitere Modifikationen vornehmen. Die KI dient dabei nicht nur als Antwortlieferant, sondern als aktiver Bestandteil eines kreativen,

mehrstufigen Arbeitsprozesses. Dies war im Datensatz meistens während der Bilderstellung mittels ChatGPT zu erkennen.

Ein typisches Beispiel findet sich in gekürzter Form im folgenden Auszug:

[2 tippt und liest die Prompt-Zeile von ChatGPT: „Erstelle ein Bild des Sonnenkönigs.“]

2: „Ich sage jetzt absichtlich Sonnenkönig, damit er sich zwar an Ludwig XIV orientiert, aber vielleicht ein bisschen kreativ wird.“

[...] Die KI erstellt ein Bild.

1: „Der schaut genauso aus wie der andere, hat das gleiche Gewand an. Sollte er nicht anders aussehen?“

2: „Machen wir Vollbart und Kurzhaar. Und Militäruniform. Machen wir aus Ludwig dem 14. einen Militärdiktator aus dem 19. Jahrhundert.“

[...] Die KI erzeugt eine neue Variante.

2: „So, was sollen wir jetzt machen? Stelle Ludwig XIV in ...“

1: „In heutiger Zeit dar?“

[2 tippt: „Stelle Ludwig den XIV im Stil des 21. Jahrhunderts mit vergleichbaren Herrschaftsinsignien dar, sei KREATIV“]

[...] KI erzeugt erneut ein Bild, beide sind zufrieden]

[2 beginnt ins Antwortfeld der Aufgabe 2.1 zu tippen.]

1: „Schreib vielleicht zuerst, dass wir ihn modernisiert haben.“

[2 schreibt: „Wir haben ihn modernisiert: Herrschaftssymbole wurden beibehalten (...) jedoch im modernen Stil, mit Jeans, T-Shirt, Goldkette, Rolex und Nikes dargestellt.“]

– Paar 1–2, Zeile 94–138

Diese Sequenz zeigt einen mehrstufigen, kreativen Bearbeitungsprozess. Die Schüler:innen nutzen zunächst einen allgemeinen Prompt, bewerten die Antwort kritisch und initiieren gezielt weitere Varianten. Im Verlauf werden unterschiedliche historische Szenarien erprobt. Schließlich fließen die KI-Ergebnisse in die eigene schriftliche Antwort ein, die wiederum eigenständig strukturiert und sprachlich angepasst wird.

6. Qualitative Ergebnisse

Im folgenden Kapitel werden die qualitativen Ergebnisse der vorliegenden Arbeit näher erläutert.

6.1. Deskriptive Verteilung der Codes

Um die genaue Verteilung der einzelnen Codes auf die teilnehmenden Paare sichtbar zu machen, wurde in MaxQDA eine Code-Dokumenten-Matrix (Abbildung 5) erstellt. Diese visualisiert auftretende Häufigkeiten von Codes und erleichtert die weitere Analyse.

Codesystem	Paar_1_2_markiert	Paar_3_4_markiert	Paar_5_6_markiert	Paar_7_8_markiert	Paar_9_10_markiert	Paar_11_12_markiert	Paar_13_14_markiert	SUMME
▼ Inhaltliche Codes								0
▼ KI-Strategien								0
▼ A: Prompting-Strategien								0
A1: Lineares Prompting	5	6	9	6	7	6	8	47
A2: Exploratives Prompting	3	3	5	2		2	1	16
A3: Iteratives Prompt-Repair	7	1	11	2	8	4	5	38
▼ B: Interaktionsformen mit KI								0
B1: KI als Informationsquelle	4	7	9	6	5	7	8	46
B2: KI als Partner	2		2		1			5
▼ C: Metakognitive Strategien								0
C1: Planung des Vorgehens	7	8	17	6	8	5	11	62
C2: Bewertung	10	5	11		3	3	3	35
C3: Strategiewechsel	2					2	1	5
▼ D: Soziale Koordination								0
D1: Rollenverteilung	1	1	1	1		2	1	7
D2: Aushandlung, Entscheid	6	6	15	5	5	4	7	48
▼ E: Technische Interaktion								0
E1: Bedienung	14	15	25	13	14	17	20	118
▼ FUER-Kompetenzen								0
F1: Sachkompetenz	11	8	1	6				26
F2: Methodenkompetenz	14	11	4	2	2	1		34
F3: Fragekompetenz	6	3		2				11
F4: Orientierungskompetenz	5	6	2	3		1		17
▼ Intensitätsniveau								0
S0: Keine KI-Nutzung für Beantw	4	6	1	3	2			16
S1: Oberflächliche Nutzung	2	4	6	6	6	6	5	35
S2: Funktionale Nutzung	4	3	1	2	1	2	1	14
S3: Reflexive, iterative Nutzung	1	1	2			1		5
Σ SUMME	108	94	122	65	62	63	71	585

Abbildung 5: MaxQDA Code-Dokumenten-Matrix

Insgesamt wurden 585 inhaltliche Kodierungen vergeben. Der Rest entfiel auf Kodierungen der Sprecher:innen, die für die weitere Analyse irrelevant sind. Besonders stark vertreten sind die Kategorien Prompting-Strategien, metakognitive Strategien und technische Interaktion, während die Kategorien soziale Koordination, FUER-Kompetenzen und Intensitätsniveaus geringere Häufigkeiten aufweisen. Bei der Kategorie der Intensitätsniveaus ist dies dem Modus der Kodierung geschuldet, da, wie bereits erwähnt, oft mehrere Sinneinheiten gemeinsam mit einem Intensitätsniveau kodiert wurden. Im Folgenden werden die Ergebnisse der einzelnen Kategorien kurz beschrieben. Eine ausführlichere Darstellung mit Definitionen, Indikatoren und Ankerbeispielen befindet sich im Methodikkapitel.

Kategorie A, „Prompting-Strategien“, erfasst, wie Schüler:innen ihre Eingaben an ChatGPT formulieren. Im Datensatz treten alle drei Unterkategorien A1–A3 mehrfach auf. A1, „lineares Prompting“, und A3, „iteratives Prompt-Repair“, kommen in allen Paaren vor, während A2, „exploratives Prompting“, zwar ebenfalls in fast allen Paaren genutzt wird, insgesamt aber deutlich seltener festgestellt werden konnte.

Kategorie B, „Interaktionsformen mit KI“, erfasst die Rolle, die ChatGPT im Bearbeitungsprozess einnimmt. B1, „KI als Informationsquelle“, ist in allen Paaren ausgeprägt und zeigt eine relativ gleichmäßige Verteilung ohne extreme Ausreißer. B2, „KI als Partner“, ist seltener und nur in einigen wenigen Paaren sichtbar.

Kategorie C, „Metakognitive Strategien“, erfasst Äußerungen zur Planung, Bewertung und Anpassung des Vorgehens. Sie weist hinter der Kategorie E, der technischen Interaktion, insgesamt die häufigsten Werte im gesamten Codesystem auf. Die Matrix zeigt eine deutliche Dominanz der Unterkategorien C1, „Planung des Vorgehens“, und C2, „Bewertung“, die fast über alle Paare hinweg stark vertreten sind. C3, „Strategiewechsel“, kommt dagegen nur vereinzelt in drei Paaren vor.

Kategorie D, „Soziale Koordination“, beschreibt, wie die Paare ihre Zusammenarbeit organisieren. Insgesamt tritt sie seltener auf als andere Kategorien. D1, „Rollenverteilung“, erscheint in den meisten Paaren nur einmal, D2, „Aushandlung/Entscheidung“, wurde dagegen in allen Paaren mit moderater Häufigkeit beobachtet, wobei das Paar 5–6 einen deutlichen positiven Ausreißer darstellt.

Kategorie E, „Technische Interaktion“, umfasst die operative Ebene der Arbeit mit ChatGPT, also die Bedienung, das Kopieren/Einfügen und Absenden von Prompts. Der Code E1, „Bedienung“, weist eine der höchsten Gesamthäufigkeiten des Codesystems auf. Sämtliche Paare zeigen hier starke Aktivität, mit besonders hohen Dichten in den Paaren 5–6 und 13–14.

Kategorie F, „FUER-Kompetenzen“, verortet Äußerungen im Rahmen des FUER-Modells des historischen Denkens. Insgesamt zeigen sich diese in moderater Häufigkeit, allerdings mit deutlichen Unterschieden zwischen den Unterkategorien und Paaren. Sachkompetenz, F1 (26 Kodierungen), findet sich vor allem in den Paaren 1–2, 3–4 und 7–8. Methodenkompetenz, F2 (34 Kodierungen), ist die am häufigsten codierte Kompetenz und tritt vor allem in den Paaren 1–2, 3–4 und 5–6 auf. Fragekompetenz, F3 (11 Kodierungen), verteilt sich auf nur drei Paare, Orientierungskompetenz, F4 (17 Kodierungen), ist etwas breiter gestreut, bleibt insgesamt aber ebenfalls seltener.

Kategorie S, „Intensitätsniveau“, beschreibt Tiefe und Reflexionsgrad der KI-Nutzung. Die Kodierungen verteilen sich unterschiedlich über die Paare hinweg. Am stärksten sind die Stufen S1, „oberflächliche Nutzung“ (35 Kodierungen), und S2, „funktionale Nutzung“ (14 Kodierungen), ausgeprägt. S0, „keine KI-Nutzung“ (16 Kodierungen), tritt nur punktuell in einigen Paaren auf, während die höchste Stufe S3, „reflexive, iterative Nutzung“, mit insgesamt 5 Kodierungen recht rar bleibt und nur in vier Paaren vorkommt. Es zeigt sich damit eher eine Konzentration auf die niedrigeren und mittleren Stufen, während reflexive Nutzungsmuster nur selten beobachtet wurden.

6.2. Grundlegende Muster im qualitativen Material

Auf Basis der Code-Dokumenten-Matrix (Abbildung 5) lassen sich mehrere Muster erkennen, die sich insbesondere auf das Zusammenspiel von KI-Strategien, FUER-Kompetenzen und Intensitätsniveaus beziehen.

Auffällig ist, dass die Nutzung von KI-Strategien und das Auftreten von FUER-Kompetenzen nicht direkt proportional zueinanderstehen. Paare mit einer sehr intensiven Nutzung von ChatGPT während der Bearbeitung kompetenzorientierter Aufgaben zeigen nicht notwendigerweise gleichzeitig eine höhere Anzahl an FUER-Kompetenz-Kodierungen. So weist etwa das Paar 5–6 eine sehr hohe Dichte an KI-Strategien auf, während die entsprechenden FUER-Kodierungen allerdings gering bleiben. Ein ähnliches Bild zeigt sich bei Paar 13–14, welches zahlreiche KI-bezogene Codes, jedoch keine FUER-Kompetenzen aufweist. Umgekehrt verfügt das Paar 1–2 über die höchsten Ausprägungen historischer Kompetenzen im gesamten Datensatz, befindet sich allerdings hinsichtlich ihres Einsatzes von KI-Strategien höchstens im Mittelfeld.

Ein weiteres Muster ergibt sich aus dem Vergleich der Intensitätsniveaus mit den FUER-Kompetenzen. Paare, die bei manchen Aufgaben völlig auf die KI-Nutzung verzichten (S0), zeigen eine höhere Anzahl an FUER-Kodierungen. Dies trifft insbesondere auf die Paare 1–2 und 3–4 zu, bei denen viele S0-Codierungen gemeinsam mit der höchsten Anzahl an FUER-Codes im gesamten Datensatz auftreten. Dagegen finden sich bei Paar 13–14 keine Sequenzen ohne KI-Nutzung und gleichzeitig keine FUER-Codes.

Die wenigen Fälle reflexiver, iterativer Nutzung (S3) streuen über den gesamten Datensatz und treten sowohl in Paaren mit hohen als auch in Paaren mit niedrigen FUER-Kompetenzwerten auf. Insgesamt zeigt sich dementsprechend kein konsistentes Verteilungsmuster. Reflexive KI-Nutzung erscheint im vorliegenden Material unabhängig von den Kompetenzniveaus der Paare.

Diese Beobachtungen legen nahe, dass nicht die generelle Häufigkeit der KI-Nutzung entscheidend ist, sondern die Art und Weise, wie ChatGPT in den Bearbeitungsprozess eingebettet wird. Um diese unterschiedlichen Umgangsweisen im Datenmaterial systematisch zu fassen, wird im folgenden Abschnitt eine Typenbildung vorgenommen.

6.3. Typenbildung nach Kuckartz im Umgang mit ChatGPT

Im Folgenden wird zur Beantwortung der zentralen Forschungsfrage der vorliegenden Arbeit die Typenbildung nach Kuckartz angewandt und erläutert.¹⁰⁵ Dies wird schematisch in Abbildung 6 gezeigt.

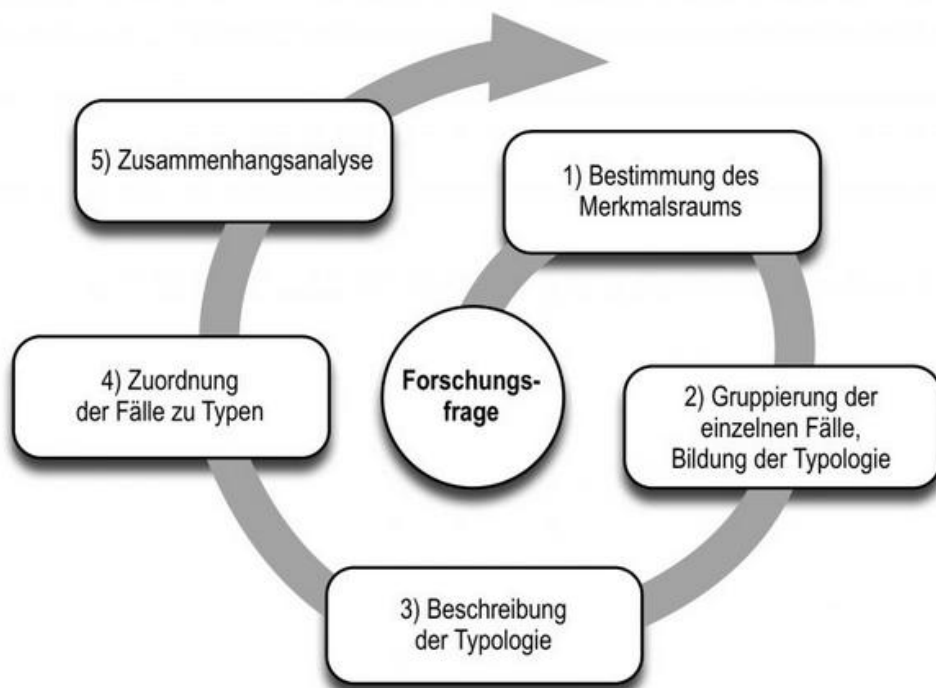


Abbildung 6: Darstellung der Phasen der Typenbildung.

Quelle: Udo Kuckartz, *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, (Weinheim 2016) 148.

¹⁰⁵ Kuckartz, *Qualitative Inhaltsanalyse*.

6.3.1. Merkmalsraum und Typenbildung

Die Typenbildung wird benötigt, um den qualitativen Datensatz in einer Weise zu strukturieren, die den Umgang der Schüler:innen mit der KI systematisch vergleichbar und damit die zentrale Forschungsfrage beantwortbar macht, welche zum besseren Verständnis hier nochmals erwähnt wird:

„Welche Strategien lassen sich beim Umgang mit ChatGPT identifizieren, die Schüler:innen nutzen, um kompetenzorientierte, in den Anforderungsbereichen I–III verortete geschichtsdidaktische Aufgaben zu bearbeiten?“

Während die Code-Dokumenten-Matrix aus MaxQDA erste Hinweise auf unterschiedliche Nutzungsmuster liefert, ermöglicht erst die Typenbildung nach Kuckartz eine Ordnung in Strategietypen, die charakteristische Kombinationen aus KI-Nutzung, metakognitivem Vorgehen und historischen Kompetenzen abbilden. Dies stellt einen zentralen Punkt im Mixed-Methods-Design der vorliegenden Arbeit dar, da es dadurch möglich wird, die qualitativen Daten der Interaktionsprotokolle mit den quantitativen Daten des Fragebogens zu verbinden und auf dieser Basis tiefergehende Schlussfolgerungen zu ziehen.

Der Fokus der Typologie liegt dabei vor allem auf jenen Merkmalen, die unmittelbar erkennen lassen, wie Schüler:innen ChatGPT in einem kompetenzorientierten Arbeitskontext einsetzen. Dazu zählen das Ausmaß und die Art der KI-Interaktion (Codes A, B), die Einbindung eigener Denkprozesse (C), das Auftreten von FUER-Kompetenzen (F1–4) sowie die Intensität der Nutzung von KI (S0–3). Diese Merkmale bilden den in der Abbildung 6 auftretenden mehrdimensionalen Merkmalsraum, in welchem sich die Typologie bewegt.¹⁰⁶ Die im Kodiersystem ebenfalls enthaltenen Kategorien bezüglich der sozialen Koordination (D) und der technischen Interaktion mit der Oberfläche (E) wurden für die eigentliche Typenbildung nicht herangezogen. Sie wurden zwar zur Beschreibung des Bearbeitungskontextes kodiert, zeigten im Datensatz jedoch nur eine geringe Differenzierungskraft zwischen den Paaren und liefern im Vergleich zu den anderen Kategorien keinen ausreichenden zusätzlichen Erkenntnisgewinn in Bezug auf die zentrale Forschungsfrage.

Auf Grundlage des oben definierten Merkmalsraums wurde im nächsten Schritt die Typologie gebildet. Dabei wurden Fälle anhand charakteristischer Kombinationen von Merkmalsausprägungen gruppiert. Ziel war es, Paare zu identifizieren, die in ihrem Umgang mit der KI vergleichbare Muster zeigen und sich dabei deutlich von anderen unterscheiden.

¹⁰⁶ Kuckartz, Qualitative Inhaltsanalyse, 153.

Die Bildung der Typologie stützte sich primär auf die in Abbildung 5 dargestellte Code-Dokumenten-Matrix in MaxQDA, welche die Häufigkeiten sämtlicher relevanter Kategorien pro Paar darstellt. Dadurch konnten einige Paare identifiziert werden, die ähnliche Profile in Bezug auf KI-Nutzung, metakognitive Steuerung, Intensitätsniveaus und FUER-Kompetenzen aufwiesen. Dies wurde anschließend durch erneutes Sichten der originalen Transkripte kontrolliert, um die Plausibilität der Gruppierungen zu prüfen.

Durch dieses Vorgehen konnten drei deutlich voneinander unterscheidbare Strategietypen identifiziert werden:

- 1. Reflexiv-selektive KI-Nutzung**
- 2. Outputorientierte KI-Nutzung**
- 3. Situativ-explorative KI-Nutzung**

Diese Typen sind jeweils durch eine spezifische Konfiguration der vier zentralen Dimensionen „KI-Strategien“, „Metakognition“, „FUER-Kompetenzen“ sowie „Intensitätsniveau“ geprägt. Sie bilden die Grundlage für die anschließende detaillierte Beschreibung sowie Fallzuordnung. In Tabelle 1 auf der nächsten Seite werden diese Typen überblicksmäßig hinsichtlich ihrer Merkmale dargestellt. Im darauffolgenden Abschnitt werden sie genauer erläutert.

Tabelle 1: Überblick über Strategietypen im Umgang mit ChatGPT und typische Merkmalsdimensionen

Strategietyp	KI-Nutzung (A- und B-Codes)	Metakognition (C-Codes)	FUER-Kompetenzen (F-Codes)	Intensitätsniveaus (S-Codes)	Beispiel-Paare
Reflexiv-selektive KI-Nutzung	Selektiver, bewusst gesteuerter Einsatz von ChatGPT. Dazwischen wiederkehrende Phasen ohne KI-Nutzung, in denen Aufgaben eigenständig bearbeitet werden. KI wird dabei eher als Werkzeug zur Erweiterung eigener Überlegungen eingesetzt, statt als reines „Lexikon“.	Hohe Ausprägung von C1 („Planung des Vorgehens“) und C2 („Bewertung“) quer über die Bearbeitungszeit. Ausgeprägte metakognitive Steuerung des KI-Einsatzes, selten tritt dabei zusätzlich C3 („Strategiewechsel“) auf.	Höchste Dichte an FUER-Kodierungen im Datensatz, insbesondere F1 („Sachkompetenz“) und F2 („Methodenkompetenz“), aber auch F3 („Fragekompetenz“) und F4 („Orientierungskompetenz“).	Mischung aus S0 („Phasen ohne KI-Nutzung“) und S2 („funktionale Nutzung“), punktuell S3 („reflexive, iterative Nutzung“). Insgesamt zeigt sich ein moderater, bewusster KI-Einsatz.	Vor allem Paar 1–2 und 3–4
Outputorientierte KI-Nutzung	Sehr intensive und stark outputorientierte Nutzung von ChatGPT. Aufgabenstellungen werden dabei häufig ohne Anpassung übernommen und an die KI delegiert. Die Antworten werden weitgehend unverändert in das Antwortfeld kopiert.	Metakognitive Äußerungen beziehen sich überwiegend auf Effizienz und Bedienung, beispielsweise beim Kürzen von KI-Antworten, und weniger auf die inhaltliche Qualität. C1 und C2 sind dabei schwach ausgeprägt, C3 tritt kaum auf.	FUER-Kompetenzen treten kaum bis gar nicht auf. Historische Denkprozesse werden weitgehend an die KI ausgelagert und nur selten eigenständig sichtbar.	Überwiegend tritt S1 („oberflächliche Nutzung“) auf. Kaum S0 und nur vereinzelt S2, S3 tritt praktisch nicht auf.	Besonders deutlich bei Paar 5–6 und 13–14
Situativ-explorative KI-Nutzung	Wechselhaftes, wenig konsistentes Muster, in welchem sich Phasen eigenständiger Bearbeitung mit spontanen, oft oberflächlichen KI-Anfragen abwechseln. Die Delegation an ChatGPT erfolgt situativ.	C1 und C2 sind punktuell gut erkennbar, werden aber nicht über längere Strecken aufrechterhalten. C3 erfolgt durch Wechsel zwischen eigenständigem Arbeiten und KI-Delegation.	FUER-Kompetenzen, vor allem F1 und F2, vereinzelt F3 und F4, treten punktuell auf, werden aber selten vertieft. Das historische Denken bleibt häufig fragmentarisch.	Heterogenes Muster aus S0, S1 und S2. S3 zeigt sich nur vereinzelt, womit die Intensität insgesamt als stark schwankend einzustufen ist.	Vor allem Paar 7–8, 9–10 sowie teilweise 11–12

6.3.2. Beschreibung der Strategietypen und Zuordnung der Fälle

Im Folgenden werden die drei im Datenmaterial gefundenen Strategietypen beschrieben und durch Beispiele aus dem Datenmaterial illustriert.

Strategietyp 1, die „Reflexiv-selektive KI-Nutzung“, ist dadurch gekennzeichnet, dass ChatGPT selektiv und bewusst eingesetzt wird, während immer wieder Phasen erfolgen, in denen eigenständig und ohne KI-Unterstützung nachgedacht oder gearbeitet wird. Dieser Typus tritt vor allem in den Paaren 1-2 und 3-4 auf. In den Transkripten sind wiederholt Passagen ohne KI-Nutzung (S0) zu finden, in denen selbstständig Aufgaben interpretiert, Fragen formuliert und historische Bedeutungen erschlossen werden. In genau diesen Sequenzen treten außerdem die meisten FUER-Kodierungen des gesamten Datensatzes auf. Beispielhaft ist hier Aufgabe 1.3, in welcher das Paar 1-2 eigenständig und ohne Hilfe von ChatGPT zwei Fragen zur Selbstdarstellung Ludwigs XIV. entwickelt und dabei mit der Analogie eines „Midlife-Crisis-Vaters mit gelbem Porsche“ arbeitet. Ein weiteres Beispiel ist die folgende typische S0-Kodierung bei Paar 1-2:

[Lesen Aufgabe 1.4]

1: „Was ist Propaganda?“

2: „Wenn du es unter das Volk bringst.“

1: „Frag mal Chat-GPT.“

2: „Wieso, das kann ich auch so beantworten.“

1: „Okay.“

- Paar 1–2, Zeile 61–66

Gleichzeitig wird die KI gezielt für explorative oder kreativitätsbezogene Teilaufgaben eingesetzt. In Aufgabe 2.1 experimentieren die Paare dieses Typs beispielsweise mit verschiedenen Bildprompts und bewerten die erzeugten Bilder kritisch, bevor die Ergebnisse in eigenen Worten im Antwortfeld zusammengefasst werden. Diese Kombination aus gezieltem KI-Einsatz und metakognitiver Steuerung unterscheidet die „Reflexiv-selektive KI-Nutzung“ deutlich von den beiden anderen Strategietypen. Die KI wird dabei mehr als Werkzeug zur Erweiterung des eigenen Denkens benutzt als nur als Lexikon. Weiters wird nicht der komplette Denkprozess an ChatGPT delegiert, sondern es treten Phasen auf, in denen völlig auf diesbezügliche Hilfe verzichtet wird. Die hohe Anzahl an FUER-Kodierungen bei gleichzeitig moderater, reflektierter KI-Nutzung macht diesen Strategietyp zu einem guten Beispiel für kompetenzorientierten Umgang mit ChatGPT im vorliegenden Datensatz.

Strategietyp 2, die „Outputorientierte KI-Nutzung“, kennzeichnet sich durch eine über weite Teile der Aufgabenbearbeitung hinweg sehr intensive KI-Nutzung. Paare dieses Typs geben Fragen oder Aufgabenstellungen meistens ohne Veränderung und unreflektiert an die KI weiter, lesen die Antworten kurz quer und übernehmen diese dann ganz oder weitgehend unverändert in das Antwortfeld. Eigenständige kognitive Leistungen treten dabei eher in den Hintergrund. Dieser Typ findet sich besonders ausgeprägt bei den Paaren 5–6 und 13–14. In vielen dieser Sequenzen wird die KI bereits verwendet, bevor die Aufgabe überhaupt gelesen oder verstanden wurde. Ein Beispiel aus Paar 13–14:

[Ohne die Frage 1.3 gelesen zu haben]

14: „Ich kopiers wieder rein.“

[Frage 1.3 wird ohne Abänderung in die Prompt-Zeile kopiert und abgeschickt. 14 liest erst jetzt die Frage, dann die Antwort von ChatGPT.]

14: „Sehr stark, sind zwei gute Fragen.“

[Die Antwort von ChatGPT wird unverändert in das Antwortfeld 1.3 kopiert.]

- Paar 13-14, Zeile 31–36

Ähnliche Muster zeigen sich auch in anderen Paaren. Aufgabenstellungen oder ganze Fragenfelder werden ohne Anpassung an ChatGPT übergeben und die generierten Antworten anschließend direkt in das Antwortfeld auf dem Aufgabenbogen kopiert. Die KI fungiert hier an erster Stelle als Antwortgenerator, als Lexikon, welches den größten Teil der Bearbeitung übernimmt. Der Denkprozess, welchen die kompetenzorientierte Aufgabe eigentlich anstoßen wollte, wird weitgehend an die KI ausgelagert. Die Schüler:innen steuern zwar die technische Seite des Prozesses, doch die inhaltliche Auseinandersetzung mit dem historischen Kontext bleibt deutlich begrenzt.

In diesem Strategietyp treten FUER-Kompetenzen nur selten oder sogar gar nicht auf. Sie werden kaum eigenständig entfaltet, sondern höchstens indirekt durch die KI-Antworten berührt. Beispielsweise ließ eines der Paare im Datensatz ChatGPT so stark in die Aufgabe einfließen, dass im gesamten Transkript kein einziger FUER-Code gefunden werden konnte. Metakognitive Prozesse beziehen sich überwiegend auf die Bedienung und Effizienz der KI, beispielsweise beim Nachdenken darüber, wie man die KI-Antwort am besten kürzen könnte, und weniger auf die Qualität der inhaltlichen Ergebnisse. Die Schüler:innen übernehmen historische Denkprozesse hier nicht selbst, wodurch kaum rekonstruktive, reflexive oder interpretative Leistungen sichtbar werden.

Strategietyp 3, die „Situativ-explorative KI-Nutzung“, beschreibt schließlich jene Paare, deren Umgang mit ChatGPT zwischen eigenständigen Denkphasen und spontanen, oft oberflächlichen KI-Anfragen schwankt, ohne dass sich ein konsistentes Muster erkennen lässt. Dieser Typ ist vor allem bei den Paaren 7–8, 9–10 und teilweise 11–12 ausgeprägt.

„Explorativ“ meint hier nicht die Promptkategorie (A2), sondern ein wechselhaftes Vorgehen zwischen Eigenleistung und KI-Delegation. Charakteristisch ist, dass die Schüler:innen stellenweise historische Kompetenzen zeigen, etwa wenn sie Aufgaben interpretieren, Fragen formulieren oder Symbole deuten. Diese Phasen bleiben jedoch kurz oder werden unmittelbar von einer KI-Delegation abgebrochen.

Ein typisches Beispiel aus Paar 7–8 zeigt zunächst selbstständige Bearbeitung ohne KI (S0) und sichtbare FUER-Kompetenzen:

8: *„Wie hättest du es formuliert?“*

7: *„Ich überlege grade.“*

8: *„Was machte ihn so besonders? Dass er allein herrschen konnte?“*

[7 und 8 formulieren gemeinsam die Frage zur historischen Selbstdarstellung und tragen sie selbst ins Antwortfeld ein.]

– Paar 7-8, Zeile 33–40

Hier sind Sachkompetenz, Rekonstruktion und Aushandlung klar erkennbar. Doch diese Arbeitsweise wird über weite Teile nicht beibehalten, sondern wechselt sich mit einem oberflächlichen Modus ab, bei welchem Aufgaben ohne Auseinandersetzung an die KI delegiert werden.

[7 kopiert die Aufgabe 3.2 unverändert in die Prompt-Zeile von ChatGPT und schickt ab.]

8: *„Jetzt erzählt er dir einen Roman.“*

[7 schreibt „Kurzfassung“ in die Prompt-Zeile von ChatGPT und schickt ab.]

[7 kopiert die Antwort von ChatGPT unverändert und ohne sie zu lesen in das Antwortfeld von Frage 3.2]

– Paar 7-8, Zeile 108–111

Diese Inkonsistenz ist auch im kreativen Aufgabenbereich vorhanden, wo die Paare zwar oft kurzfristig eigene Ideen entwickeln, diese aber rasch durch KI-Generierung ersetzen und sich nicht recht lange mit der Aufgabe beschäftigen.

Auch FUER-Kompetenzen erscheinen punktuell, werden jedoch nicht vertieft. Schüler:innen dieses Typs arbeiten zeitweise zwar historisch kompetent, zeigen jedoch insgesamt ein

inkonsistentes Leistungsmuster, welches stark davon abhängt, wie sich die Paare in der jeweiligen Situation entscheiden vorzugehen. Dieser Typ unterscheidet sich von Typ 1 mit seiner stabileren Kompetenzentwicklung und von Typ 2 mit stärkerer KI-Dominanz, indem er keinen festen strategischen Stil entwickelt, sondern zwischen beiden Polen schwankt.

Um die gewonnenen Strategietypen weiter einzuordnen und mit quantitativen Merkmalen der Stichprobe in Beziehung zu setzen, werden im folgenden Kapitel die Ergebnisse der Fragebogenerhebung dargestellt.

7. Quantitative Ergebnisse

Im folgenden Teil der Arbeit werden die Antworten der 14 teilnehmenden Personen quantitativ analysiert. Ziel der Analyse ist es, zentrale Konstrukte im Umgang mit ChatGPT, beispielsweise Wissen, Nutzungsverhalten, Einstellungen, Selbstwirksamkeit und Vertrauen zu erfassen, deren Zusammenhänge zu explorieren und Anknüpfungspunkte für zukünftige Forschung aufzuzeigen.

Grundlage bildet dabei der aus SoSci-Survey in SPSS (Version 31) überführte Datensatz. Aufbauend auf den im Methodikkapitel beschriebenen Skalen wurden zunächst deskriptive Werte und die Reliabilitätsmaße (Cronbachs Alpha) berechnet. Dies wurde um eine Faktorenanalyse zur Prüfung der Faktorladungen ergänzt. Anschließend wurden die Beziehungen zwischen den gefundenen Konstrukten mithilfe von Spearman-Rangkorrelationen analysiert, da überwiegend ordinalskalierte Daten vorlagen und die Größe der Stichprobe den Einsatz nichtparametrischer Verfahren nahelegt.

7.1. Deskriptive Statistik

Insgesamt nahmen 14 Schüler:innen des WIFI St. Pölten an der Erhebung teil. Die Geschlechterverteilung war vollständig ausgeglichen mit jeweils 7 weiblichen und 7 männlichen Teilnehmer:innen. Angaben zu „divers“, „keine Angabe“ oder „nicht beantwortet“ lagen nicht vor.

Das Alter lag zwischen 16 und 21 Jahren ($M = 18,71$; $SD = 1,64$). Auch wenn es sich dabei formal um Personen aus der Erwachsenenbildung handelt, sind die Teilnehmenden hinsichtlich Alter und Lernzielen gut mit älteren Schüler:innen der Sekundarstufe II vergleichbar, welche bereits Erfahrungen im schulischen und außerschulischen Umgang mit digitalen Technologien gesammelt haben.

Im Anschluss werden die deskriptiven Kennwerte der erhobenen Skalen, welche im Methodikkapitel genauer aufgeschlüsselt werden, dargestellt. Für jede davon werden Mittelwert, Median, Standardabweichung, Minimum, Maximum und Spannweite berichtet. Alle Skalen basieren auf 14 gültigen Fällen.

Die Skala „Wissen über KI/ChatGPT“ setzte sich aus fünf dichotomisierten Items zusammen, mit denen der Wissensstand der Schüler:innen zu KI und ChatGPT erfasst wurde (0–5 Punkte). Die Teilnehmenden erreichten einen Mittelwert von $M = 0,93$ mit einer Standardabweichung von $SD = 0,92$. Die Werte lagen zwischen 0 und 3, der Modus bei 1 Punkt. Dies deutet auf einen eher geringen Wissenstand bezüglich KI unter den Teilnehmenden hin.

Der niedrige Score könnte dabei auch durch allgemeine Unsicherheit der Schüler:innen und die gewählte Dichotomisierung der Skala verstärkt worden sein.

Die Skala „Diversität in der Nutzung von ChatGPT“ erfasste mit sieben Items, wie vielfältig die Lernenden ChatGPT für schulische Tätigkeiten einsetzen (Likert 1–5). Die Werte reichten von 1 (einseitige Nutzung) bis 5 (sehr vielfältige Nutzung). Der Mittelwert betrug $M = 3,06$, die Standardabweichung $SD = 1,31$ (Spannweite: 4 Punkte). Damit zeigte sich eine mittlere Diversität der Nutzung.

Die Skala „Positive Einstellung gegenüber ChatGPT“ umfasste drei Items (Likert 1–5), welche die Aufgeschlossenheit gegenüber ChatGPT abbilden. Die Schüler:innen erzielten einen Mittelwert von $M = 3,12$, bei einer Standardabweichung von $SD = 0,78$. Die Werte lagen zwischen 1,67 und 4,00 (Spannweite: 2,33 Punkte). Die Ergebnisse sprechen für eine eher positive Haltung gegenüber ChatGPT.

Die Skala „Negative Einstellung gegenüber ChatGPT“ bestand aus drei Items (Likert 1–5), welche wahrgenommene Risiken oder negative Aspekte von ChatGPT abbilden. Der Mittelwert lag bei $M = 3,31$, die Standardabweichung bei $SD = 0,80$. Die Werte reichten von 1,33 bis 4,33 (Spannweite: 3 Punkte). Damit zeigen sich die Schüler:innen moderat bewusst gegenüber den möglichen Risiken.

Die Skala „Selbstwirksamkeit im Umgang mit ChatGPT“, welche das Vertrauen in die eigenen Fähigkeiten bezüglich des KI-Einsatzes abbildet, umfasste sieben Items (Likert 1–5). Die Schüler:innen erreichten einen Mittelwert von $M = 3,68$ bei einer Standardabweichung von $SD = 1,19$. Die Werte lagen zwischen 1,57 und 5,00 (Spannweite: 3,43 Punkte). Die Ergebnisse zeigen eine insgesamt eher höhere wahrgenommene Eigenkompetenz im Umgang mit ChatGPT.

Die Skala „Vertrauen in ChatGPT“ bestand aus vier Items (Likert 1–5). Der Mittelwert betrug $M = 2,23$, die Standardabweichung $SD = 0,91$. Die Werte lagen zwischen 1,00 und 3,75 (Spannweite: 2,75). Das Vertrauen in ChatGPT fällt damit insgesamt eher niedrig aus.

Die Variable „Nutzungsfrequenz von ChatGPT“ wurde ordinal auf einer 7-stufigen Skala erfasst (1 = nie, 7 = mehrmals täglich). Die häufigste Kategorie war „mehrmals pro Woche“ (Modus = 5). Die Werte reichten von 1 bis 7 (Spannweite: 6 Punkte). Vier Personen nutzten ChatGPT nie, eine Person mehrmals im Monat, fünf mehrmals pro Woche und vier mehrmals täglich. Damit ergab sich eine bimodale Verteilung, die eine deutliche Trennung zwischen Nicht-Nutzenden und Vielnutzenden erkennen lässt.

7.2. Korrelationstabelle der erhobenen Skalen

Zur Untersuchung der Zusammenhänge zwischen den sieben erhobenen Konstrukten wurden paarweise Spearman-Rangkorrelationen berechnet. Tabelle 2 zeigt die Korrelationsmatrix der einzelnen Skalen und ist farblich kodiert. Grüne Felder markieren dabei positive, rote Felder negative Korrelationen. Mit einem Stern gekennzeichnete Koeffizienten sind bei einem zweiseitigen Signifikanzniveau von $p < .05$ statistisch signifikant. Bei mit zwei bzw. drei Sternen gekennzeichneten Werten liegt dieses Niveau auf $p < .01$ bzw. $p < .001$.

Tabelle 2: „Heatmap“ der Spearman-Korrelationen der untersuchten Skalen

	Wissen über KI/ChatGPT	Diversität in der Nutzung von ChatGPT	Positive Einstellung gegenüber ChatGPT	Negative Einstellung gegenüber ChatGPT	Selbstwirksamkeit im Umgang mit ChatGPT	Vertrauen in ChatGPT	Nutzungsfrequenz von ChatGPT
Wissen über KI/ChatGPT	--						
Diversität in der Nutzung von ChatGPT	,097	--					
Positive Einstellung gegenüber ChatGPT	,142	,415	--				
Negative Einstellung gegenüber ChatGPT	,170	,468	,092	--			
Selbstwirksamkeit im Umgang mit ChatGPT	,245	,787***	,534*	,131	--		
Vertrauen in ChatGPT	-,281	,481	,585*	,183	,532*	--	
Nutzungsfrequenz von ChatGPT	,297	,830***	,344	,342	,827***	,364	--

Die Korrelationsanalyse zeigt mehrere inhaltlich relevante Zusammenhänge zwischen den erhobenen Konstrukten. Aufgrund der kleinen Stichprobe ($N = 14$) müssen die Ergebnisse allerdings vorsichtig interpretiert werden. Dennoch lassen sich einige Muster erkennen. Im Folgenden wird die Korrelationstabelle anhand der Effektstärke Rho und der statistischen Signifikanz p näher erläutert.

Für die Skala „Wissen über KI/ChatGPT“ zeigten sich keine signifikanten Zusammenhänge mit den anderen Skalen. Lediglich ein leicht negativer Trend zum Vertrauen in ChatGPT ($Rho = -.28$) und ein leicht positiver Trend zur Nutzungsfrequenz ($Rho = .30$) waren erkennbar. Dies könnte darauf hindeuten, dass das fachliche Wissen der Teilnehmenden über ChatGPT weitgehend unabhängig von der tatsächlichen Nutzung, von Einstellungen oder vom erlebten Umgang mit der KI ist.

Die Skala „Diversität der Nutzung“ zeigte mehrere deutliche Zusammenhänge. Besonders stark korrelierte sie mit der Nutzungsfrequenz ($Rho = .83$, $p < .001$), was darauf hindeuten

könnte, dass häufigere Nutzung auch mit einer breiteren Anwendung der KI einhergeht. Eine ebenso starke und signifikante Korrelation ergab sich bei der Selbstwirksamkeit im Umgang mit ChatGPT ($Rho = .79, p < .001$). Teilnehmende, die ChatGPT in verschiedenen Anwendungsbereichen einsetzen, schätzen ihre eigene Kompetenz im Umgang mit der KI höher ein. Weitere Zusammenhänge mit Einstellungen oder Vertrauen waren zwar leicht positiv, erreichten jedoch keine statistische Signifikanz.

Die Skala „Positive Einstellung gegenüber ChatGPT“ stand in einem signifikanten Zusammenhang mit der Selbstwirksamkeit ($Rho = .53, p = .049$). Personen aus der Stichprobe, die ChatGPT positiv bewerten, schätzen also auch ihre individuellen Fähigkeiten, mit der KI umzugehen, höher ein. Ebenso zeigte sich eine signifikant positive Beziehung zum Vertrauen in ChatGPT ($Rho = .58, p = .028$). Weitere positive, aber nicht signifikante Zusammenhänge bestanden zur Diversität der Nutzung, zur Nutzungsfrequenz sowie zur Wissensskala. Dies könnte darauf hindeuten, dass eine positive Einstellung der KI gegenüber mit dem eigenen Kompetenzzempfinden und dem generellen Vertrauen in ChatGPT verbunden ist.

Die Skala „Negative Einstellung gegenüber ChatGPT“ stand in keinem signifikanten Zusammenhang mit den anderen Skalen, trotz leicht positiver Trends zu Diversität der Nutzung ($Rho = .47$) und zur Nutzungsfrequenz ($Rho = .34$). Damit bleibt die Skala in dieser Stichprobe als ein weitgehend eigenständiges Konstrukt, welches nur schwach mit anderen Merkmalen zusammenhängt.

Die Skala „Selbstwirksamkeit im Umgang mit ChatGPT“ zeigte zwei sehr starke und signifikante Zusammenhänge, einerseits mit der Diversität der Nutzung ($Rho = .79, p < .001$) und andererseits mit der Nutzungsfrequenz ($Rho = .83, p < .001$). Teilnehmende, die ChatGPT regelmäßig und in verschiedenen Situationen verwenden, schätzen sich auch deutlich kompetenter im Umgang damit ein. Zudem bestand ein moderater, aber signifikanter Zusammenhang zur positiven Einstellung gegenüber ChatGPT ($Rho = .53, p = .049$) sowie ein ebenfalls knapp signifikanter positiver Zusammenhang zum Vertrauen in ChatGPT ($Rho = .53, p = .050$). Keine statistische Bedeutung zeigten die leicht positiven Korrelationen zum Wissen über KI/ChatGPT und zur negativen Einstellung.

Die Skala „Vertrauen in ChatGPT“ stand in einem signifikanten Zusammenhang mit der positiven Einstellung gegenüber ChatGPT ($Rho = .58, p = .028$) sowie mit der Selbstwirksamkeit ($Rho = .53, p = .050$). Teilnehmende, die ChatGPT positiver bewerten und sich im Umgang damit kompetent fühlen, bringen der KI tendenziell mehr Vertrauen entgegen. Alle übrigen Korrelationen waren nicht signifikant und überwiegend nur schwach ausgeprägt.

Das Vertrauen scheint somit in der Stichprobe vor allem mit Einstellungen und dem eigenen Kompetenzzempfinden verknüpft.

Die Variable „Nutzungsfrequenz von ChatGPT“ zeigte zwei sehr starke und signifikante Zusammenhänge. Einerseits mit der Diversität der Nutzung ($Rho = .83, p < .001$) und andererseits mit der Selbstwirksamkeit ($Rho = .83, p < .001$). Teilnehmende, die ChatGPT häufiger einsetzen, nutzen es nicht nur breiter, sondern schätzen auch ihre eigene Kompetenz deutlich höher ein. Weitere Zusammenhänge waren zwar überwiegend positiv, aber statistisch nicht bedeutsam.

7.3. Explorative Zusammenhänge zwischen Strategietypen und Skalen

Um das volle Potenzial des Mixed-Methods-Ansatzes der vorliegenden Arbeit auszuschöpfen, wurden Zusammenhänge zwischen den qualitativ erarbeiteten Strategietypen und den erhobenen quantitativen Skalen explorativ geprüft. Dazu wurden Dummy-Variablen erstellt, die die Zugehörigkeit der einzelnen Paare zu den jeweiligen Strategietypen 1–3 abbilden (0 = keine Zugehörigkeit, 1 = Zugehörigkeit). Diese Dummy-Variablen wurden anschließend mittels Spearman-Rangkorrelation mit den oben erläuterten Skalen zu Wissen, Nutzung, Einstellungen, Vertrauen und Selbstwirksamkeit in Beziehung gesetzt. Die vollständige Korrelationsmatrix ist in Tabelle 3 dargestellt. Mit einem Stern gekennzeichnete Koeffizienten sind bei einem zweiseitigen Signifikanzniveau von $p < .05$ statistisch signifikant. Bei mit zwei bzw. drei Sternen gekennzeichneten Werten liegt dieses Niveau auf $p < .01$ bzw. $p < .001$. Aufgrund der sehr kleinen Fallzahl sind die Zusammenhänge als explorative Hinweise zu interpretieren.

Für den Strategietyp 1 „reflexiv-selektive KI-Nutzung“ zeigten sich signifikant negative Zusammenhänge mit der Nutzungsdiversität ($Rho = -.67, p = .009$), der Selbstwirksamkeit ($Rho = -.73, p = .003$) und der Nutzungsfrequenz ($Rho = -.60, p = .025$). Damit zeigt sich, dass ein gezielter, eher zurückhaltender KI-Einsatz in der Stichprobe vor allem bei Paaren auftrat, welche die KI seltener und weniger vielfältig nutzen.

Strategietyp 2, die „outputorientierte KI-Nutzung“, korrelierte positiv mit der positiven Einstellung gegenüber ChatGPT ($Rho = .59, p = .027$) sowie mit der Selbstwirksamkeit ($Rho = .69, p = .006$). Bei der Nutzungsfrequenz wurde ein positiver, jedoch nicht signifikanter Trend festgestellt ($Rho = .45, p = .105$). Ein outputorientierter Modus mit hoher KI-Beteiligung findet in der Stichprobe also eher bei Paaren mit hoher Selbstwirksamkeit und positiver Haltung gegenüber ChatGPT.

Für Typ 3, die „situativ-explorative KI-Nutzung“, ergaben sich keine bedeutsamen Zusammenhänge mit den erhobenen Skalen. Unklar ist, ob dies auf eine tatsächlich fehlende Korrelation oder auf die begrenzte statistische Power der Analyse zurückzuführen ist.

Tabelle 3: „Heatmap“ der Spearman-Korrelationen der untersuchten Strategietypen mit den untersuchten Skalen

Spearman-Korrelationen der untersuchten Strategietypen mit den untersuchten Skalen (Heatmap)										
	Nutzungsfrequenz von ChatGPT	Wissen über KI/ChatGPT	Diversität in der Nutzung von ChatGPT	Positive Einstellung gegenüber ChatGPT	Negative Einstellung gegenüber ChatGPT	Selbstwirksamkeit im Umgang mit ChatGPT	Vertrauen in ChatGPT	Strategietyp 1	Strategietyp 2	Strategietyp 3
Nutzungsfrequenz von ChatGPT	--									
Wissen über KI/ChatGPT	,297	--								
Diversität in der Nutzung von ChatGPT	,830***	,097	--							
Positive Einstellung gegenüber ChatGPT	,344	,142	,415	--						
Negative Einstellung gegenüber ChatGPT	,342	,170	,468	,092	--					
Selbstwirksamkeit im Umgang mit ChatGPT	,827***	,245	,787***	,534*	,131	--				
Vertrauen in ChatGPT	,364	-,281	,481	,585*	,183	,532*	--			
Strategietyp 1	-,596*	,105	-,669**	-,357	-,139	-,731**	-,416	--		
Strategietyp 2	,452	,335	,413	,588*	-,178	,692**	,317	-,400	--	
Strategietyp 3	,131	-,401	,234	-,211	,290	,036	,090	-,548*	-,548*	--

8. Diskussion

Im Folgenden werden die zentralen Befunde der qualitativen und quantitativen Analyse kurz zusammengefasst, um danach Schlussfolgerungen daraus zu ziehen.

Auf qualitativer Ebene zeigen sich zwischen den Paaren deutliche Unterschiede im Umgang mit ChatGPT. Während grundlegende Prompting-Strategien bei allen vorzufinden waren, unterschieden sich die einzelnen Fälle stark hinsichtlich der Tiefe ihres Vorgehens. Eine zentrale Beobachtung an dieser Stelle ist, dass die Häufigkeit KI-bezogener Strategien nicht direkt proportional zur Anzahl gezeigter historischer FUER-Kompetenzen ist. Bei Paaren mit intensiver KI-Nutzung wurden nur geringe oder gar keine FUER-Kompetenzen codiert, während bei Paaren, welche die KI weniger stark nutzten, mehr FUER-Codes vergeben wurden. Weiters zeigte sich, dass Sequenzen ohne KI-Nutzung häufig gemeinsam mit FUER-Kodierungen auftraten. Reflexive KI-Nutzung wurde generell nur selten und quer über den Datensatz hinweg festgestellt.

Auf quantitativer Ebene zeigt die Stichprobe ein geringes Wissensniveau über KI und ChatGPT, während die Einstellungen gegenüber dem Chatbot heterogen ausfallen und das Vertrauen ihm gegenüber eher niedrig ist. Im Gegensatz dazu war die Selbstwirksamkeit etwas höher und die Diversität moderat ausgeprägt. Die Korrelationsanalyse nach Spearman ergab dabei mehrere deutliche Zusammenhänge. Eine hohe Nutzungsfrequenz geht mit einer breiteren Anwendungsvielfalt und gleichzeitig hoher Selbstwirksamkeit einher. Ebenso steht eine positive Einstellung zu ChatGPT in signifikantem Zusammenhang mit höherer Selbstwirksamkeit sowie einem höheren Vertrauen gegenüber dem Chatbot. Das Wissen bezüglich KI hingegen korreliert mit keiner der erhobenen Variablen in einer nennenswerten Weise. Die Verteilung der Stichprobe zeigt Differenzen zwischen Personen, welche ChatGPT nicht oder nur selten, und Personen, welche ChatGPT oft nutzen, hin. Dieser Unterschied spiegelt sich dabei sowohl in der Nutzungsfrequenz als auch der Selbstwirksamkeit wider.

8.1. Strategietypen im Kontext der quantitativen Skalen

Die im qualitativen Teil herausgearbeiteten drei Strategietypen zeigen in der explorativen Zusammenhangsanalyse unterschiedliche Korrelationen mit den im Fragebogen erhobenen Skalen. Aufgrund der geringen Fallzahl lassen sich daraus keine generalisierbaren Aussagen ableiten, die Muster liefern jedoch wichtige Hinweise zur Interpretation der Strategien.

Typ 1, „reflexiv-selektive KI-Nutzung“, geht in der vorliegenden Stichprobe mit einer geringeren Nutzung von ChatGPT einher und benutzt KI generell weniger vielfältig. Weiters

ist dieser Strategietyp mit einer niedrigeren Selbstwirksamkeit in Bezug auf die Benutzung des Chatbots verbunden. Paare, welche im qualitativen Material besonders oft FUER-Codes erhalten haben, sind dementsprechend weniger selbstsicher beim Umgang mit der KI und benutzen diese auch weniger oft. In der vorliegenden Arbeit zeigten sich historische Kompetenzen vor allem, wenn die KI selektiv und mehr als Ergänzung zum eigenen Denkprozess genutzt wurde, und zentrale Schritte der Aufgabenbearbeitung ohne die KI erfolgten.

Dem Strategietyp 2, „outputorientierte KI-Nutzung“, angehörige Personen im Datensatz weisen im Gegensatz dazu eine höhere Selbstwirksamkeitserwartung und eine insgesamt positivere Einstellung gegenüber ChatGPT auf. Dies spiegelt sich in den im qualitativen Teil beschriebenen Sequenzen wider, in denen der Chatbot häufig und sehr selbstverständlich eingesetzt wird, wodurch große Teile des Denkprozesses an die KI delegiert werden. Die Zusammenhangsanalyse deutet darauf hin, dass in der vorliegenden Stichprobe eine hohe subjektive Sicherheit und eine positive Haltung gegenüber ChatGPT, dem Werkzeug, mit dem gearbeitet werden sollte, nicht automatisch mit einer tieferen Auseinandersetzung mit der Aufgabe selbst einhergehen muss. Stattdessen scheint ein stark ausgeprägtes Vertrauen in die eigenen Fähigkeiten bezüglich der KI-Nutzung dazu beizutragen, dass der Chatbot völlig in den Vordergrund tritt und dabei eigenständige historische Denkprozesse zurückdrängt.

Typ 3, die „situativ-explorative KI-Nutzung“, zeigt bei den Skalenwerten dagegen kein klares Muster. Dies passt gut zum qualitativen Ergebnis, dass die Paare dieses Strategietyps zwischen eigenständigen Denkphasen und Delegation an die KI schwanken und keinen stabilen Stil bezüglich ChatGPT entwickeln. Eine mögliche Deutung wäre daher, dass sich in diesem Typ Schüler:innen finden, die noch keinen gefestigten Umgang mit ChatGPT bezüglich solcher Aufgaben entwickelt haben und deswegen mehrere Wege ausprobieren. Gleichzeitig lässt sich aufgrund der geringen Fallzahl und der sehr heterogenen Verteilung auch annehmen, dass der Mischtyp generell unterschiedliche Hintergründe vereint. Beispielsweise verschiedene Vorerfahrungen mit KI, unterschiedliche fachliche Voraussetzungen oder situative Einflüsse im Erhebungskontext. Vermutlich reichte die statistische Power der vorliegenden Studie mit N=14 nicht aus, um hier trennschärfere Muster herauszuarbeiten.

8.2. Reflexion des Forschungsdesigns

Im Unterschied zu den im Kapitel der Geschichtsdidaktik vorgestellten Studien von Mierwald und Santamaría-Velasco et al., die vor allem die Interaktionen mit simulierten historischen Akteur:innen untersuchen, fokussiert sich die vorliegende Arbeit auf Bildquellen, KI-

Bildgenerierung und die Analyse eines aktuellen Werbeplakats. Dadurch werden spezielle Aspekte des strategischen KI-Einsatzes sichtbar, beispielsweise der Umgang mit Prompt-Optimierung und der Einsatz von ChatGPT als Unterstützung bei der Bildinterpretation. Zusätzlich unterscheidet sich die vorliegende Arbeit von den genannten Studien durch die Verknüpfung der beobachteten Strategien mit dem FUER-Modell und den Anforderungsbereichen I–III sowie durch den Mixed-Methods-Ansatz. Ein weiterer Unterschied besteht außerdem darin, dass die Studien von Kindenberg und Mierwald auch die Qualität der von der KI erzeugten Produkte untersuchen. Die vorliegende Arbeit verzichtet bewusst auf eine systematische Bewertung des Outputs, da diese den Rahmen sprengen würde. Stattdessen liegt der Fokus auf dem Umgang der Schüler:innen mit ChatGPT selbst, also auf den Strategien, metakognitiven Äußerungen und historischen Kompetenzen, die dabei sichtbar werden.

Für die zentrale Forschungsfrage bedeuten diese Ergebnisse, dass der kompetenzorientierte Umgang mit ChatGPT im vorliegenden Datensatz nicht durch eine hohe Routine oder ein besonders ausgeprägtes Vertrauen in die KI als Werkzeug gekennzeichnet ist, sondern eher durch die Fähigkeit, die Nutzung der KI gezielt zu begrenzen und in den Phasen dazwischen eigenes historisches Denken einzusetzen. Der Chatbot soll nicht ersetzen, sondern ergänzend eingesetzt werden. Auf dieser Grundlage lassen sich Implikationen für kompetenzorientierten Unterricht formulieren, welche im nächsten Kapitel näher erläutert werden.

9. Fazit

9.1. Beantwortung der Forschungsfragen

Zentraler Punkt der vorliegenden Arbeit war die Frage, wie Schüler:innen ChatGPT einsetzen, wenn sie kompetenzorientierte, in den Anforderungsbereichen I–III verortete geschichtsdidaktische Aufgaben bearbeiten und welche Strategien sich dabei identifizieren lassen. Ergänzend wurde dabei untersucht, inwiefern diese Strategietypen mit quantitativ erfassten Merkmalen wie Nutzungsfrequenz, Diversität der Nutzung, Wissen über KI, Selbstwirksamkeit und Einstellungen gegenüber ChatGPT zusammenhängen. Grundlegendes Ziel war es, erste empirische Einsichten in die Rolle von KI-Nutzung für historisches Lernen im schulischen Kontext zu gewinnen. Damit lässt sich die vorliegende Arbeit in die noch recht junge Forschung zu Chatbots in Bildungskontexten einreihen.¹⁰⁷

Die grundlegende Forschungsfrage nach Strategietypen im Umgang mit ChatGPT lässt sich auf Basis der qualitativen Analyse klar beantworten. In der untersuchten Stichprobe konnten drei unterscheidbare Typen definiert werden. Typ 1, die „reflexiv-selektive KI-Nutzung“, beschreibt Paare, die ChatGPT eher gezielt und bewusst einsetzen, dazwischen immer wieder längere Phasen ohne KI-Nutzung einlegen und in diesen eigenständig Aufgaben interpretieren, Fragen formulieren, diskutieren und aushandeln. Genau in diesen Sequenzen traten die meisten FUER-Kodierungen des Datensatzes auf.

Typ 2, die „outputorientierte KI-Nutzung“, hingegen zeichnet sich dadurch aus, dass Aufgabenstellungen und Fragen weitgehend unverändert an ChatGPT weitergegeben und die zurückkommenden Antworten nahezu ungeprüft übernommen werden. Historische Denkprozesse werden weitgehend an die KI ausgelagert und FUER-Kompetenzen bleiben selten.

Der letzte Typ, die „situativ-explorative KI-Nutzung“, vereint dagegen Paare, die zwischen eigenständigen Denkphasen und spontanen, intensiven KI-Phasen schwanken, ohne einen stabilen Bearbeitungsstil auszubilden. FUER-Kompetenzen zeigen sich punktuell, werden aber häufig durch starke KI-Nutzung unterbrochen. Das Hauptmerkmal dieses Typs ist damit die Inkonsistenz.

Die vorliegende Arbeit zeigt damit auf, dass sich bei der Bearbeitung von kompetenzorientierten Aufgaben durch Schüler:innen mittels ChatGPT kein einheitliches

¹⁰⁷ Yirga Yayeh *Munaye*, Wasyihun *Admass*, Yenework *Belayneh*, Atinkut *Molla*, Mekete *Asmare*, ChatGPT in Education: A Systematic Review on Opportunities, Challenges, and Future Directions. In: *Algorithms* 18 (2025) 352, doi.org/10.3390/a18060352.

Nutzungsmuster beobachten lässt, sondern mindestens drei qualitativ unterschiedliche Strategietypen, nämlich ein eher KI-begrenzender, ein KI-dominanter und ein inkonsistenter Umgang mit KI. Historische Kompetenzen im Sinne des FUER-Modells wurden dabei vor allem dort sichtbar, wo die KI nur als ergänzendes Werkzeug in einen primär eigenständigen Denkprozess eingebettet wurde und nicht die gesamte Bearbeitung dominierte.

Die ergänzenden Forschungsfragen, welche nach möglichen Zusammenhängen zwischen den definierten Strategietypen und verschiedenen quantitativen Merkmalen fragen, lassen sich aufgrund der geringen Stichprobengröße nur sehr begrenzt beantworten. Dennoch liefert die Analyse einige Hinweise. In der vorliegenden Stichprobe zeigte sich, dass Strategietyp 1 mit einer geringeren Nutzungsfrequenz sowie Anwendungsvielfalt und mit einer niedrigeren Selbstwirksamkeit im Umgang mit ChatGPT korreliert war. Dagegen wiesen Schüler:innen des Strategietyps 2 eine höhere Selbstwirksamkeit und eine positivere Einstellung gegenüber ChatGPT auf. Für den Strategietyp 3 ergab sich kein klares Skalenprofil.

Ein genereller Zusammenhang zwischen verschiedenen Strategietypen und dem im Fragebogen abgefragten Wissen über KI konnte nicht festgestellt werden. Diese Variable spielte im Datensatz offenbar diesbezüglich keine wichtige Rolle. Die Ergebnisse deuten dementsprechend eher darauf hin, dass die verschiedenen Typen in der vorliegenden Stichprobe vor allem mit den Einstellungen, der Selbstwirksamkeit und den Nutzungsmustern bezüglich KI verknüpft sind, weniger mit dem Wissensstand darüber.

Für die grundlegende Forschungsfrage lässt sich festhalten, dass ein kompetenzorientierter Umgang mit ChatGPT im untersuchten Unterrichtskontext nicht durch eine hohe Routine, ein ausgeprägtes Vertrauen in KI als Werkzeug oder eine höhere Nutzungsintensität gekennzeichnet ist. Stattdessen scheint es wichtiger zu sein, die KI-Nutzung gezielt zu begrenzen und sie in Phasen einzubetten, welche von eigenständigem, historischem Denken geprägt sind, um eine Entfaltung historischer Kompetenzen in KI-Settings zu ermöglichen. Die vorliegende Arbeit macht damit sichtbar, dass es weniger darauf ankommt, wie viel mit KI gearbeitet wird, sondern vielmehr, wie diese Nutzung konkret gestaltet und in FUER-bezogenes historisches Lernen eingebettet wird.

Die Ergebnisse reihen sich damit in die bisherige geschichtsdidaktische Forschung zu ChatGPT ein, welche im Kapitel zur Geschichtsdidaktik vorgestellt wurde.

9.2. Implikationen für Unterricht und Lehrer:innenbildung

Die vorliegende Arbeit macht deutlich, dass der Einsatz von ChatGPT zur Lösung von klassisch formulierten, kompetenzorientierten Aufgaben der Anforderungsbereiche I-III nicht automatisch zu einem Mehrwert hinsichtlich der Förderung historischen Denkens und somit der Qualität des Unterrichts führt. Vielmehr zeigt sich, dass Schüler:innen ChatGPT sehr unterschiedlich in den Lernprozess einbinden. Aus einer geschichtsdidaktischen Perspektive ergeben sich diesbezüglich verschiedene Implikationen für die Gestaltung des Unterrichts, in denen KI-Tools bewusst in Verbindung mit FUER-Konzepten gebracht werden.

9.2.1. Didaktische Chancen und Risiken von ChatGPT im Geschichtsunterricht

In Hinblick auf kompetenzorientierte Modelle historischen Lernens, die historisches Denken als ein Zusammenspiel aus Sach-, Methoden-, Frage- und Orientierungskompetenz begreifen, lässt sich ChatGPT prinzipiell als ein nützliches Werkzeug sehen. Es bietet für die Schüler:innen Zugang zu einem „Meer“ aus Informationen, welches praktisch immer und überall sofort verfügbar ist und sich an die Nutzer:innen anpasst. Es bietet Formulierungshilfen und eröffnet völlig andere Perspektiven als bisherige Tools. In dieser Rolle kann der Chatbot beispielsweise den Schüler:innen dabei helfen, plötzlich aufkommende Interessensfragen zu einem bestimmten Thema zu beantworten, Fragen an eine historische Quelle zu formulieren, die den Schüler:innen gegeben wird, alternative Deutungen sichtbar zu machen oder erste Strukturierungen für Antworttexte zu entwickeln. Eine Nutzung in dieser Weise ist anschlussfähig an die Idee eines konstruktiven Lernens mittels digitaler Medien, welche nicht ersetzend, sondern anregend in historische Lernprozesse eingebunden werden. Diese möglichen Chancen werden ebenso in verschiedenen Studien aufgezeigt. So heben Munaye et al. (2025) die Wichtigkeit der Art der didaktischen Einbettung von ChatGPT für die Effektivität der KI für den Unterricht hervor und sprechen sich dafür aus, dass zukünftige Studien nach dem optimalen Weg diesbezüglich suchen sollen.¹⁰⁸

Es gibt jedoch auch eine Kehrseite der Medaille, welche in der vorliegenden Arbeit ebenfalls ersichtlich wurde. Einerseits sind die Einschränkungen von LLMs, welche im dazugehörigen Kapitel bereits ausgiebig diskutiert wurden, zu bedenken. Andererseits zeigt der im qualitativen Material aufgekommene Strategietyp 2, was passieren kann, wenn die KI in einer völlig freien Weise in den Unterricht eingebracht wird. ChatGPT kann schnell eine viel zu dominante Rolle im Bearbeitungsprozess von Aufgaben einnehmen, in welcher Fragestellungen völlig ungefiltert delegiert und zurückgegebene Antworten ungeprüft übernommen werden. In dieser

¹⁰⁸ Munaye, ChatGPT in Education.

Hinsicht ist auch die bereits vorgestellte Studie von Mierwald (2024) relevant, in welcher ChatGPT selbst bei sehr grundlegenden historischen Fragen falsche Rückmeldungen gab, was gerade bei einer unreflektierten Übernahme seitens der Schüler:innen zu Problemen führen kann.¹⁰⁹ In der vorliegenden Arbeit wurden die Antworten der Paare auf dem Antwortbogen nicht in genauerem Maße analysiert, da dies den Rahmen gesprengt hätte. Aus der Perspektive geschichtsdidaktischer Kompetenzmodelle ist schlussendlich die Gefahr zu bedenken, dass der Chatbot die zentralen Aufgaben historischen Denkens völlig übernimmt und die Schüler:innen sich „zurücklehnen“. Es entsteht zwar der Anschein hoher Leistungsfähigkeit, da der Output für die jeweilige Schulstufe oft dennoch gut ist, jedoch ohne tatsächliche Sach-, Fragen-, Methoden- oder Orientierungskompetenz aufzubauen. Diese Problematiken müssen daher sowohl im Unterricht als auch in der Lehrer:innenbildung ausdrücklich thematisiert und didaktisch eingeplant werden.

9.2.2. Leitlinien für kompetenzorientierten KI-Einsatz und Lehrer:innenbildung

Der Strategietyp 1 macht deutlich, wie eine sinnvolle Nutzung von ChatGPT aussehen könnte, welche mit sichtbarer Entfaltung von FUER-Kompetenzen einhergeht. Phasen von eigener Rekonstruktion, Diskussion und sozialer Aushandlung wechseln sich mit KI-Interaktionen ab, werden von diesen aber nicht zu sehr überlagert. Für die Unterrichtsgestaltung lassen sich daraus eventuelle didaktische Leitlinien ableiten:

Erstens können Lehrpersonen beim Umgang mit KI Phasen einplanen, in denen Schüler:innen Aufgaben zunächst ohne technische Unterstützung bearbeiten oder sich zumindest darüber Gedanken machen müssen. Beispielsweise bei einer Bildanalyse erste Interpretationen selbst formulieren, erste Hypothesen zu Sachverhalten selbst entwickeln oder eigene Fragen an eine Quelle sammeln. Im Anschluss kann ChatGPT als Ergänzung hinzugenommen werden, beispielsweise um die Hypothesen zu testen, bei der Formulierung zu helfen oder zusätzliche Fragen zu generieren.

Zweitens könnten Aufgaben so gestaltet werden, dass Antworten der KI nicht als Endprodukt akzeptiert werden, sondern von den Schüler:innen begründet, überprüft oder korrigiert werden müssen. Beispielsweise könnte mit anderen, zusätzlichen Quellen und Werkzeugen gearbeitet werden oder die Unterschiede zwischen den eigenen, im Voraus gemachten Hypothesen und den Vorschlägen des Chatbots herausgearbeitet werden.

¹⁰⁹ Mierwald, Chatting about the Past.

Drittens legt die Rekonstruktion der drei Strategietypen nahe, dass sich Schüler:innen teilweise sehr unterschiedlich bei der Verwendung von ChatGPT in Verbindung mit Arbeitsaufträgen verhalten. Genau diesen Umstand könnte die Lehrperson thematisieren, indem sie im Unterricht selbst Raum für die Reflexion über die von den Schüler:innen gewählten Strategien schafft. Dies kann durch kurze Metagespräche, Reflexionsbögen oder den Vergleich unterschiedlicher Vorgehensweisen geschehen und eventuell dazu beitragen, vom delegierenden hin zu einem ergänzenden, kontrollierenden Einsatz von KI zu gelangen.

Diese Leitlinien sollen nicht nur die Entwicklung historischer Kompetenzen im Unterricht unterstützen, sondern greifen auch die Überlegungen der „AI-Literacy-Forschung“ auf. So betonen Zhong und Liu (2025), dass der kompetente Umgang mit KI über die bloße Bedienung selbst hinausgeht und auch die Fähigkeiten umfasst, Antworten der KI kritisch einzuordnen, die Grenzen des Werkzeugs zu erkennen und dessen Nutzung zu reflektieren.¹¹⁰ Die Ergebnisse der vorliegenden Arbeit zeigen in dieser Hinsicht beispielsweise, dass eine hohe Selbstwirksamkeit, eine positive Einstellung gegenüber ChatGPT sowie eine hohe Nutzungsfrequenz sowie -variabilität in der Stichprobe allein nicht ausreichen, um einen solchen reflektierten Umgang sicherzustellen. Stattdessen muss die KI in ein geeignetes Unterrichtsgeschehen eingebettet werden, um die Entfaltung von zusätzlichen Kompetenzen zu ermöglichen.

Die erfassten Strategietypen sind nicht nur für den Unterricht selbst und damit involvierte Lehrpersonen interessant, sondern werfen auch auf der übergeordneten Ebene der Lehrer:innenbildung Fragen auf. Wenn Schüler:innen ChatGPT tatsächlich in so unterschiedlicher Weise nutzen, müssen Lehrpersonen in der Lage sein, dies zu erkennen und didaktisch darauf zu reagieren. Einerseits gehört dazu ein grundlegendes Verständnis von der Funktion und den Grenzen der KI, andererseits aber auch fachspezifisches Wissen darüber, wie sich die FUER-Kompetenzen in einem KI-Lernsetting fördern lassen. In dieser Hinsicht betonen auch Zhong und Liu (2025), wie wichtig die Schulung von Lehrpersonen bezüglich AI-Literacy ist.¹¹¹

Dies bedeutet, dass KI nicht nur als technisches „Randthema“, sondern als grundlegender Bestandteil der fachdidaktischen Ausbildung verstanden werden sollte. Da ChatGPT in Zukunft vermutlich von den meisten Schüler:innen genutzt werden wird, müssen zukünftige Lehrpersonen der Geschichte und anderer Fächer diesbezüglich besser ausgebildet werden. Es müssen Gelegenheiten geschaffen werden, Unterrichtssequenzen zu erproben, in denen

¹¹⁰ Baichang Zhong, Xiaofan Liu, Evaluating AI literacy of secondary students: Framework and scale development. In: Computers & Education 227 (2025) 105230, doi.org/10.1016/j.compedu.2024.105230.

¹¹¹ Zhong, Liu, Evaluating AI literacy of secondary students, 15–18.

ChatGPT genutzt wird, und diese Nutzung muss anschließend reflektiert werden. Die in der vorliegenden Arbeit identifizierten Strategietypen können dabei als heuristische Vorlagen dienen, um typische Nutzungsmuster von Schüler:innen zu thematisieren und mögliche Stolpersteine, beispielsweise die völlige Delegation des Denkprozesses, frühzeitig zu erkennen und zu umgehen.

9.3. Limitationen

Die Untersuchung in der vorliegenden Arbeit ist durch verschiedene Limitationen gekennzeichnet, welche abschließend an dieser Stelle erwähnt werden.

Eine der zentralsten Begrenzungen der Aussagekraft stellt die Größe und der Kontext der Stichprobe dar. Der Datensatz wurde auf Basis von 14 Schüler:innen eines Kurses des WIFI St. Pölten im Alter von 16 bis 21 erstellt. Dementsprechend handelt es sich dabei um eine recht kleine Gruppe aus einem sehr spezifischen Bildungskontext. Weiters war die Teilnahme an der Untersuchung freiwillig, wodurch mit einem gewissen Selbstselektionsbias zu rechnen ist. In dieser Hinsicht sind die Ergebnisse nicht ohne Weiteres auf andere Schulformen oder Schulstufen übertragbar.

Auch wenn beim Erstellen des Fragebogens darauf geachtet wurde, verschiedene Themen und Kompetenzen zu bedienen, so war das Feld der angesprochenen Inhalte dennoch recht begrenzt. Andere Unterrichtsthemen mit anderen Methoden könnten zu völlig abweichenden Strategien hinsichtlich der Benutzung von ChatGPT führen.

Die qualitativen Daten der vorliegenden Arbeit wurden mittels der „Think-Aloud“-Methode gesammelt. Dabei wurde ein spezieller Ansatz gewählt, bei dem zwei Schüler:innen gleichzeitig gemeinsam die Aufgaben bearbeiteten. Durch die Wahl dieser Methode konnten zwar sehr reichhaltige Dialoge generiert werden, jedoch wird dabei das Arbeitsverhalten der involvierten Teilnehmenden beeinflusst. Dies wurde beispielsweise durch Rollenverteilungen sichtbar, bei denen eine Person eher die Denkleistung und die andere eher die Bedienung von Tastatur und KI übernahm, wodurch die Gesamtleistung des Paares und dessen Strategien nicht ohne weiteres auf andere Settings übertragen werden können. Generell wussten die Schüler:innen, dass sie beobachtet und aufgezeichnet werden, was ihr Verhalten ebenfalls beeinflusst haben könnte.

Zusätzlich wurden die aus den Daten der Paare generierten Strategietypen als Dummy-Variablen im weiteren Verlauf mit den Antworten einzelner Personen aus dem Fragebogen verglichen. Dabei ist zu berücksichtigen, dass innerhalb eines Paares eine erhebliche Heterogenität hinsichtlich Wissen, KI-Erfahrung und Einstellungen bestehen kann. Zudem sind

die Daten innerhalb der Paare nicht unabhängig, da beide Partner:innen dieselbe Typenzugehörigkeit erhalten, was zu einer Überschätzung der statistischen Signifikanz führen kann. Die Ergebnisse sind daher mit besonderer Vorsicht zu interpretieren. Der quantitative Teil stützt sich zum größten Teil auf selbstberichtete Skalen. Mit Ausnahme der Wissensskala wurden die anderen auf Basis der subjektiven Einschätzungen der einzelnen Schüler:innen gebildet und können dementsprechend von deren tatsächlichen Verhalten abweichen. Auch wenn Reliabilität und Faktorladungen überprüft wurden, wurden die Skalen speziell auf die Bedürfnisse der vorliegenden Arbeit angepasst und sind weniger geeignet, um unverändert in einem größeren Rahmen angewandt zu werden.

Die statistischen Analysen der vorliegenden Arbeit dienen vor allem der Exploration eventuell vorhandener Muster und um Ansätze für Hypothesen zukünftiger Forschung zu generieren. Auch wenn einige Spearman-Rho-Korrelationen statistische Signifikanz erreicht haben, können daraus keine inferenzstatistischen Schlussfolgerungen gezogen werden, da aufgrund der geringen Power der Analysen Zufallseinflüsse nicht hinreichend ausgeschlossen werden können. Dies wurde an den entsprechenden Stellen explizit erwähnt. Die Ergebnisse sind kontextspezifische Einsichten in den Umgang von Schüler:innen mit ChatGPT im Geschichtsunterricht zu verstehen, ohne den Anspruch, damit repräsentative Aussagen für eine größere Population zu generieren.

9.4. Ausblick

Zuletzt werden mögliche weitere Ansatzpunkte für die zukünftige Forschung in diesem Feld diskutiert. Der naheliegendste nächste Schritt wäre eine Erweiterung des in der vorliegenden Arbeit benutzten Settings durch eine größere Stichprobe in anderen Schulformen und Jahrgängen. Damit könnte geprüft werden, ob die im Datensatz vorgefundenen Strategietypen nur auf diesen begrenzt sind, sich auch in einem breiteren Rahmen feststellen lassen oder welche anderen Typen sich noch zeigen würden, wenn insgesamt mehr Fälle verglichen werden.

Dabei festgestellte Strategietypen könnten mit Leistungsdaten verglichen werden, ein Aspekt, welcher in der vorliegenden Arbeit aus Umfangsgründen nicht vertieft wurde. Zukünftige Studien könnten bewerten, inwiefern bestimmte gezeigte Strategien tatsächlich mit einem bestimmten Niveau historischen Denkens einhergehen. An dieser Stelle sei noch einmal die Studie von Mierwald (2024) genannt, der genau diesen Aspekt mit einer ebenfalls recht kleinen Stichprobe von $N=21$ untersuchte.¹¹²

¹¹² Mierwald, Chatting about the Past.

Weiters wäre ein interdisziplinärer Ansatz wünschenswert. Die vorliegende Studie untersuchte KI-Nutzung aus einer geschichtsdidaktischen Perspektive, doch wird ChatGPT genauso in anderen Fächern genutzt. Ähnliche Studien in Deutsch, Geografie oder Biologie könnten klären, ob es fächerübergreifende Strategien gibt oder sich spezielle Ansätze herausbilden, welche sich je nach Aufgabenstellung unterscheiden.

Ein weiterer naheliegender Schritt wäre ein systematischer Vergleich zwischen ChatGPT und anderen digitalen Recherchewerkzeugen wie Wikipedia oder klassischen Suchmaschinen. Dabei könnte untersucht werden, wie sich Strategien durch die Nutzung von KI verändern. Gerade für die geschichtsdidaktische Frage nach Quellen- und Darstellungskritik wäre relevant, ob Schüler:innen bei ChatGPT andere Bewertungsstrategien anwenden als bei etablierten Informationsangeboten.

Darüber hinaus könnten die Empfehlungen der vorliegenden Arbeit gemeinsam mit Ergebnissen anderer Studien in realen Unterrichtssettings über einen längeren Zeitraum erprobt und die Effekte empirisch überprüft werden. In diesem Zusammenhang bietet auch die Perspektive der Lehrpersonen einen wichtigen Ansatz, um zu untersuchen, welche Erfahrungen und Einstellungen diese Gruppe bezüglich des Umgangs der Schüler:innen mit KI im Unterricht aufweist, was eine Herangehensweise darstellt, deren Bedeutung auch Zhong und Liu (2025) betonen.¹¹³

¹¹³ Zhong, Liu, Evaluating AI literacy of secondary students, 15–18.

10. Literaturverzeichnis

- Steffen *Albrecht*, ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen (TAB-Hintergrundpapier 26, Berlin 2023), doi.org/10.5445/IR/1000158070.
- Benedikte S. *Als*, Janne J. *Jensen*, Mikael B. *Skov*, Comparison of Think-Aloud and Constructive Interaction in Usability Testing with Children. In: Proceedings of the 2005 Conference on Interaction Design and Children (Boulder 2005) 9–16.
- Heinrich *Ammerer*, Elfriede *Windischbauer* (Hg.), Kompetenzorientierter Unterricht in Geschichte und Politischer Bildung. Diagnoseaufgaben mit Bildern (Wien 2011).
- APA, Österreichs Unternehmen sind dabei den digitalen Wandel zu verschlafen. In: derStandard.at (13.2.2024), <https://www.derstandard.at/story/3000000207231/oesterreichs-unternehmen-sind-dabei-den-digitalen-wandel-zu-verschlafen> (10.6.2025).
- APA, Forschungsrat fordert Überarbeitung der KI-Strategie. In: derStandard.at (8.10.2024), <https://www.derstandard.at/story/3000000239824/forschungsrat-fordert-ueberarbeitung-der-ki-strategie> (10.6.2025).
- Ulrich *Baumgärtner*, Wegweiser Geschichtsdidaktik. Historisches Lernen in der Schule (Paderborn 2024).
- Daniel *Bernsen*, Thomas *Spahn*, Medien und historisches Lernen. Herausforderungen und Hypes im digitalen Wandel. In: Zeitschrift für Geschichtsdidaktik 14 (2015) 191–203.
- Bundesministerium für Bildung, Wissenschaft und Forschung, Auseinandersetzung mit Künstlicher Intelligenz im Bildungssystem (Wien 2023), https://www.bmbwf.gv.at/Themen/schule/zrp/ki/ki_asbs.html (25.4.2025).
- Bundesministerium für Bildung, Wissenschaft und Forschung, Die Verwendung KI-basierter Tools beim Erstellen abschließender Arbeiten – Potenziale, Risiken und beurteilungsrelevante Aspekte. Informationen für Betreuer/innen abschließender Arbeiten an AHS und BMHS (Wien 2023), https://www.bmb.gv.at/dam/jcr:3bc6eb26-f4b1-499c-a601-675e7fd6fa0f/ki_abarb.pdf (25.4.2025).
- Bundesministerium für Bildung, Wissenschaft und Forschung, Künstliche Intelligenz – Chance für Österreichs Schulen (Wien 2024), <https://www.bmbwf.gv.at/ki> (10.6.2025).
- Bundesministerium für Bildung, Wissenschaft und Forschung, Strategie der Bundesregierung für Künstliche Intelligenz. Artificial Intelligence Mission Austria 2030 (Wien 2021),

- <https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/strategie-bundesregierung.html> (25.4.2025).
- Jürgen *Bortz*, Christof *Schuster*, Statistik für Human- und Sozialwissenschaftler (Heidelberg 2010).
- Daniel *Castro*, Michael *McLaughlin*, Eline *Chivot*, Who is winning the AI race: China, the EU or the United States? (Washington 2019), <https://www2.datainnovation.org/2019-china-eu-us-ai.pdf> (14.6.2025).
- Liang *Chen*, Guoliang *Li*, Bo *Ma*, Chenxi *Tang*, Fuhui *Okubo*, Atsushi *Shimada*, How Do Strategies for Using ChatGPT Affect Knowledge Comprehension? In: Artur M. *Olney*, Ioana-Andreea *Chounta*, Zhiyuan *Liu*, Olga C. *Santos*, Ig I. *Bittencourt* (Hg.), Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky (Communications in Computer and Information Science 2150, Cham 2024) 151–162, doi.org/10.1007/978-3-031-64315-6_12.
- Grant *Cooper*, Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. In: Journal of Science Education and Technology 32 (2023) 444–452, doi.org/10.1007/s10956-023-10039-y.
- Ana R. *Costa*, Nuno *Lima*, Catarina *Viegas*, Ana *Caldeira*, Critical minds: Enhancing education with ChatGPT. In: Cogent Education 11 (2024) 2415286, doi.org/10.1080/2331186X.2024.2415286.
- Ruiqi *Deng*, Maoli *Jiang*, Xinlu *Yu*, Yuyan *Lu*, Shasha *Liu*, Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. In: Computers & Education 227 (2025) 105224, doi.org/10.1016/j.compedu.2024.105224.
- Nicola *Döring*, Jürgen *Bortz*, Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (Berlin 2023).
- Tobias *Engelschalk*, Martin *Daumiller*, Marion *Reindl*, Markus *Dresel*, Forschungsmethoden. In: Detlef *Urhahne*, Markus *Dresel*, Frank *Fischer* (Hg.), Psychologie für den Lehrberuf (Berlin/Heidelberg 2019) 531–559.
- Europäischer Rechnungshof, Die Ambitionen der EU im Bereich der künstlichen Intelligenz: Mehr Governance und verstärkte, gezielter ausgerichtete Investitionen sind zukunftsentscheidend. Sonderbericht 2/2024 (Luxemburg 2024), <https://data.europa.eu/doi/10.2865/139576> (10.6.2025).

- Europäisches Parlament, KI-Gesetz: erste Regulierung der künstlichen Intelligenz (8.6.2023), <https://www.europarl.europa.eu/topics/de/article/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz> (10.6.2025).
- Simone *Grassini*, Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. In: *Education Sciences* 13 (2023) 692, doi.org/10.3390/educsci13070692.
- Mohammad *Halaweh*, ChatGPT in education: Strategies for responsible implementation. In: *Contemporary Educational Technology* 15 (2023) 421, doi.org/10.30935/cedtech/13036.
- Oliver *Held*, Die Rolle des Geschichtsunterrichts in Zeiten von KI. Zwischen Fakt und Fiktion: Der Einfluss von KI auf die Konstruktion und Rezeption von Geschichte. In: *Geschichte lernen* 219 (2024) 59–62.
- Abigail *Higgins*, Stephen Hawking’s final warning for humanity: AI is coming for us. In: *Vox, Future Perfect* (16.10.2018), <https://www.vox.com/future-perfect/2018/10/16/17978596/stephen-hawking-ai-climate-change-robots-future-universe-earth> (15.4.2025).
- Jan *Hodel*, Geschichtslernen mit Copy & Share. In: *hist.net – Plattform für Geschichte und digitale Medien* (11.03.2009), <https://www.hist.net/archives/2260> (21.12.2025).
- Jan *Hodel*, Historische Online-Kompetenz. In: *hist.net (Archiv), Forschung & Praxis*, o. J., <http://archiv.hist.net/forschung-praxis/historische-online-kompetenz/index.html> (21.12.2025).
- Jan *Hodel*, *Wikipedia im Geschichtsunterricht* (Frankfurt 2020).
- Jim *Holdsworth*, Matthew *Finio*, AI Examples & Business Use Cases. In: *IBM Think*, <https://www.ibm.com/think/topics/artificial-intelligence-business-use-cases> (15.4.2025).
- Jussi S. *Jauhiainen*, Antonio G. *Guerra*, Generative AI and ChatGPT in School Children’s Education: Evidence from a School Lesson. In: *Sustainability* 15 (2023) 14025, doi.org/10.3390/su151814025.
- Xuan *Ji*, Does GPT-4 have access to the internet using AI? Unveiling the truth. In: *BytePlus*, <https://www.byteplus.com/en/topic/413088> (25.8.2025).
- Charles I. *Jones*, At What Point Do We Decide AI’s Risks Outweigh Its Promise? In: *Stanford Graduate School of Business, Insights* (29.5.2024), <https://www.gsb.stanford.edu/insights/what-point-do-we-decide-ais-risks-outweigh-its-promise> (15.4.2025).

- Mohammad *Khalil*, Erol *Er*, Will ChatGPT Get You Caught? Rethinking of Plagiarism Detection. In: Panayiotis Zaphiris, Andri Ioannou (Hg.), Learning and Collaboration Technologies (Lecture Notes in Computer Science 14040, Cham 2023) 475–487, doi.org/10.1007/978-3-031-34411-4_32.
- Björn *Kindenberg*, ChatGPT-Generated and Student-Written Historical Narratives: A Comparative Analysis. In: Education Sciences 14 (2024) 530, doi.org/10.3390/educsci14050530.
- Jinhee *Kim*, Seongryeong *Yu*, Sang-Soog *Lee*, Rita *Detrick*, Students' prompt patterns and its effects in AI-assisted academic writing: Focusing on students' level of AI literacy. In: Journal of Research on Technology in Education (2025), doi.org/10.1080/15391523.2025.2456043.
- Carmen *Köhler*, Johannes *Hartig*, ChatGPT in higher education: Measurement instruments to assess student knowledge, usage, and attitude. In: Contemporary Educational Technology 16 (2024) ep528, doi.org/10.30935/cedtech/15144.
- Mareike *König*, ChatGPT und Co. in den Geschichtswissenschaften – Grundlagen, Prompts und Praxisbeispiele. In: Digital Humanities am DHIP (19.8.2024), <https://dhdhi.hypotheses.org/9197> (25.6.2025).
- Andreas *Körber*, Kompetenzorientierung in der Domäne Geschichte (Saarbrücken 2012).
- Gregor *Kramper*, USA und China hängen Europa bei der Zukunftsindustrie Künstlicher Intelligenz ab. In: stern.de (3.2.2019), <https://www.stern.de/digital/technik/kuenstliche-intelligenz---usa-und-china-haengen-europa-ab-8563640.html> (14.6.2025).
- Udo *Kuckartz*, Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (Weinheim 2016).
- Martin *Kugler*, Europa hat die Digitalisierung verschlafen. In: Die Presse, Wort der Woche (14.9.2024), <https://www.diepresse.com/18861023/europa-hat-die-digitalisierung-verschlafen> (10.6.2025).
- Victor R. *Lee*, Denise *Pope*, Sarah *Miles*, Rosalía C. *Zárate*, Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. In: Computers and Education: Artificial Intelligence 7 (2024) 100253, doi.org/10.1016/j.caeai.2024.100253.
- Spyros *Makridakis*, The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. In: Futures 90 (2017) 46–60, doi.org/10.1016/j.futures.2017.03.006.

- Mostafa *Mesgari*, Chitu *Okoli*, Mohamad *Mehdi*, Finn Årup *Nielsen*, Arto *Lanamäki*, “The Sum of All Human Knowledge”: A Systematic Review of Scholarly Research on the Content of Wikipedia. In: Journal of the Association for Information Science and Technology 66/2 (2015) 219–245, doi.org/10.1002/asi.23172.
- Marcel *Mierwald*, Chatting about the Past with Artificial Intelligence. A Case Study of Pupils’ Interaction with ChatGPT while Completing a History-Learning Task. In: Journal of Educational Media, Memory, and Society 16 (2024) 143–173, doi.org/10.3167/jemms.2024.160206.
- moinAI (knowhere GmbH), ChatGPT: Die 14 wichtigsten Fragen + Antworten. In: moinAI Chatbot-Lexikon, <https://www.moin.ai/chatbot-lexikon/chatgpt> (22.6.2025).
- Yirga Yayeh *Munaye*, Wasyihun *Admass*, Yenework *Belayneh*, Atinkut *Molla*, Mekete *Asmare*, ChatGPT in Education: A Systematic Review on Opportunities, Challenges, and Future Directions. In: Algorithms 18 (2025) 352, doi.org/10.3390/a18060352.
- Davy Tsz Kit *Ng*, Jac Ka Lok *Leung*, Samuel Kai Wah *Chu*, Maggie Shen *Qiao*, Conceptualizing AI literacy: An exploratory review. In: Computers and Education: Artificial Intelligence 2 (2021) 100041, doi.org/10.1016/j.caeai.2021.100041.
- öbv, Wie verbreitet ist Künstliche Intelligenz an Schulen? In: öbv Magazin (27.11.2023), <https://www.oebv.at/magazin/wie-verbreitet-ist-kuenstliche-intelligenz-an-schulen> (15.4.2025).
- OpenAI, GPT-4, <https://openai.com/de-DE/index/gpt-4-research/> (25.6.2025).
- OpenAI, Learning from human preferences, <https://openai.com/index/learning-from-human-preferences/> (25.6.2025).
- Francesca *Paris*, Larry *Buchanan*, 35 Ways Real People Are Using A.I. Right Now. In: The New York Times, The Upshot (14.4.2023), <https://www.nytimes.com/interactive/2023/04/14/upshot/up-ai-uses.html> (15.4.2025).
- Irmgard *Plattner*, Gemälde historischer Herrscher und Herrscherinnen als Medien der Kompetenzdiagnostik. In: Heinrich Ammerer, Elfriede Windischbauer (Hg.), Kompetenzorientierter Unterricht in Geschichte und Politischer Bildung. Diagnoseaufgaben mit Bildern (Wien 2011) 42–46.
- Rechtsinformationssystem des Bundes (RIS), Lehrpläne – allgemeinbildende höhere Schulen Anl. 1. In: Bundesrecht konsolidiert, tagesaktuelle Fassung, zuletzt geändert durch BGBl. II Nr. 204/2024,

<https://www.ris.bka.gv.at/NormDokument.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568&Anlage=1> (21.12.2025).

- Angela *Sandmann*, Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In: Dietmar *Krüger*, Ilka *Parchmann*, Horst *Schecker* (Hg.), Methoden in der naturwissenschaftsdidaktischen Forschung (Berlin/Heidelberg 2014) 179–188, doi.org/10.1007/978-3-642-37827-0_15.
- Juan *Santamaría-Velasco*, Aracelly *Núñez-Naranjo*, Ximena *Morales-Urrutia*, Critical Thinking and AI: Enhancing History Teaching through ChatGPT Simulations. In: International Journal of Innovative Research and Scientific Studies 8 (2025) 564–575, doi.org/10.53894/ijirss.v8i1.4403.
- Ghadeer *Sawalha*, Imran *Taj*, Abdulhadi *Shoufan*, Analyzing student prompts and their effect on ChatGPT's performance. In: Cogent Education 11 (2024) 2397200, doi.org/10.1080/2331186X.2024.2397200.
- Waltraud *Schreiber*, Ein Kompetenz-Strukturmodell historischen Denkens. In: Zeitschrift für Pädagogik 54 (2008) 198–212.
- Muhammad Farrukh *Shahzad*, Shuo *Xu*, Muhammad *Asif*, Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model. In: British Educational Research Journal 51 (2025) 489–513, doi.org/10.1002/berj.4084.
- Martin *Stepanek*, Österreich verspielt seine KI-Zukunft. In: derStandard.at (15.6.2023), <https://www.derstandard.at/story/3000000174849/oesterreich-verspielt-seine-ki-zukunft> (10.6.2025).
- Artur *Strzelecki*, Students' Acceptance of ChatGPT in Higher Education: An Extended Unified Theory of Acceptance and Use of Technology. In: Innovative Higher Education 49 (2024) 223–245, doi.org/10.1007/s10755-023-09686-1.
- Ulrich *Trautwein*, Christiane *Bertram*, Bodo *von Borries*, Nicola *Brauch*, Matthias *Hirsch*, Kathrin *Klausmeier*, Andreas *Körber*, Christoph *Kühberger*, Johannes *Meyer-Hamme*, Martin *Merkt*, Herbert *Neureiter*, Stephan *Schwan*, Waltraud *Schreiber*, Wolfgang *Wagner*, Monika *Waldis*, Michael *Werner*, Béatrice *Ziegler*, Andreas *Zuckowski*, Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts „Historical Thinking – Competencies in History“ (HiTCH) (Münster/New York 2017).

- Halil *Uğraş*, Murat *Uğraş*, Stamatios *Papadakis*, Michail *Kalogiannakis*, ChatGPT-Supported Education in Primary Schools: The Potential of ChatGPT for Sustainable Practices. In: *Sustainability* 16 (2024) 9855, doi.org/10.3390/su16229855.
- UNESCO, Guidance for Generative AI in Education and Research (Paris 2023), doi.org/10.54675/EWZM9535.
- Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Verordnung über künstliche Intelligenz), <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (25.4.2025).
- Chuang *Wang*, Stephen J. *Aguilar*, Joseph S. *Bankard*, Emily *Bui*, Benjamin *Nye*, Writing with AI: What College Students Learned from Utilizing ChatGPT for a Writing Assignment. In: *Education Sciences* 14 (2024) 976, doi.org/10.3390/educsci14090976.
- Jin *Wang*, Wenxiang *Fan*, The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: insights from a meta-analysis. In: *Humanities and Social Sciences Communications* 12 (2025) 621, doi.org/10.1057/s41599-025-04787-y.
- Wirtschaftskammer Österreich, Europa: Verantwortungsvolle KI (25.6.2024), <https://www.wko.at/epu/europa-verantwortungsvolle-ki> (10.6.2025).
- Ping *Zhang*, Gemma *Tur*, A systematic review of ChatGPT use in K-12 education. In: *European Journal of Education* 59 (2024) e12599, doi.org/10.1111/ejed.12599.
- Baichang *Zhong*, Xiaofan *Liu*, Evaluating AI literacy of secondary students: Framework and scale development. In: *Computers & Education* 227 (2025) 105230, doi.org/10.1016/j.compedu.2024.105230.

11. Appendix

11.1. Qualitativer Fragebogen

Aufgabe 1

Betrachte das Gemälde (Abbildung 1) „Louis XIV of France“ des Malers Hyacinthe Rigaud von 1701.

Nutze bei der Beantwortung der Fragen auf der nächsten Seite unterstützend ChatGPT.



Abbildung 7: Hyacinthe Rigaud: Louis XIV of France, online unter https://commons.wikimedia.org/wiki/Category:Louis_XIV_of_France#/media/File:Louis_XIV_of_France.jpg

1.1) Beschreiben Sie, welche Symbole der Herrschaft auf dem Bild zu erkennen sind.

1.2) Erklären Sie die historischen Rahmenbedingungen des Bildes. Geben Sie an, welche politischen Umstände für dessen Entstehung prägend waren und welcher Zweck damit verfolgt wurde.

1.3) Formulieren Sie zwei eigene Fragen zur Selbstdarstellung von Herrschern zur Zeit von Ludwig XIV.

1.4) Erörtern Sie, ob und inwiefern das Bild als Beispiel für politische Propaganda gewertet werden kann, und ziehen Sie Vergleiche zu modernen Darstellungen.

Aufgabe 2

Im Folgenden werden Sie mittels ChatGPT bzw. der integrierten Bilderstellungssoftware selbst ein Herrscher- oder Herrscherinnengemälde anfertigen.

Nutzen Sie bei der Beantwortung der Fragen unterstützend ChatGPT.

- 2.1) Erstellen Sie mit Hilfe von ChatGPT ein Bild von einem Herrscher oder einer Herrscherin deiner Wahl. Formulieren Sie dafür einen sogenannten „Prompt“ (einen Befehl an die KI), in dem Sie möglichst genau beschreiben, wie das Bild aussehen soll.

Das Ziel: Es soll wirken wie ein typisches Herrschaftsgemälde – ähnlich dem von Ludwig XIV.

Testen Sie verschiedene Formulierungen (z. B. zur Kleidung, Haltung, Hintergrund) und verfeinern Sie Ihren Prompt so lange, bis Sie mit dem Ergebnis zufrieden sind. Sprechen Sie dabei Ihre Gedanken laut aus.

- 2.2) Vergleichen Sie Ihr mit ChatGPT erstelltes Bild mit dem Gemälde aus Aufgabe 1. Gibt es Unterschiede, gibt es Gemeinsamkeiten? Diese könnten beispielsweise Posen, Symbole, Kleidung, Gesichtsausdruck, die Umgebung oder den Eindruck, den das Bild macht, umfassen.

Wie gut kann eine KI Ihrer Meinung nach den Stil historischer Herrschaftsbilder nachahmen – und wo sind ihre Grenzen?

- 2.3) Stellen Sie sich vor, Ihr KI-generiertes Herrschaftsgemälde wird in einem Geschichtslehrbuch gedruckt, ohne Hinweis darauf, dass es digital erstellt wurde. Beurteilen Sie, ob das ein Problem darstellen würde, und begründen Sie Ihre Annahme.

Aufgabe 3

Betrachten Sie das Foto (Abbildung 2) aus dem Tweet von Saul Sadka von der Onlineplattform X.

Nutzen Sie bei der Beantwortung der Fragen auf der nächsten Seite unterstützend ChatGPT. Sprechen Sie dabei Ihre Gedanken laut aus.



Abbildung 2: Saul Sadka [@Saul_Sadka], This sign is evil... In: X, ehemals Twitter (28.04.2025, 18:04), online unter: https://x.com/Saul_Sadka/status/1916886311795057089 (abgerufen am 22.06.2025).

- 3.1) Betrachten Sie das rote Werbeplakat. Beschreiben Sie genau, was auf dem Plakat zu sehen ist. Gehen Sie auf den Text, die Gestaltung (zB. Farben, Schrift, Größe), die Platzierung, etc. ein.

- 3.2) Erklären Sie, wie der abgebildete Nike-Slogan „Never again. Until next year.“ gemeint sein könnte. Gibt es diesbezüglich unterschiedliche Perspektiven? Denken Sie dabei an das historische Verständnis der Aussage „Never again“ in Europa.

11.2. Quantitativer Fragebogen

Der folgende Fragebogen wurde mit SoSci-Survey erstellt und den Teilnehmenden online zur Verfügung gestellt. Er wurde vor der quantitativen Aufgabe ausgefüllt. Im Folgenden wird er in einer formatierten Form wiedergegeben.

Einverständniserklärung

„Die Richtlinien guter ethischer Forschung sehen vor, dass sich die Teilnehmer/innen an empirischen Studien explizit und nachvollziehbar mit der Teilnahme einverstanden erklären. Freiwilligkeit. Ihre Teilnahme an dieser Untersuchung ist freiwillig. Es steht Ihnen zu jedem Zeitpunkt dieser Studie frei, Ihre Teilnahme abubrechen, ohne dass Ihnen daraus Nachteile entstehen.

Anonymität. Ihre Daten sind selbstverständlich vertraulich, werden nur in anonymisierter Form ausgewertet und nicht an Dritte weitergegeben. Demographische Angaben wie Alter oder Geschlecht lassen keinen eindeutigen Schluss auf Ihre Person zu.

Fragen. Falls Sie noch Fragen zu dieser Studie haben sollten, wenden Sie sich an den Studienleiter.

Hiermit bestätige ich, dass ich die Einverständniserklärung gelesen und verstanden habe.“

- ☐ Nein (nicht an der Studie teilnehmen)
- ☐ Ja

Demografische Angaben

Geschlecht:

„Welches Geschlecht haben Sie?“

- ☐ weiblich
- ☐ männlich
- ☐ divers
- ☐ keine Angabe

Alter:

„Wie alt sind Sie?“

Antwortformat: offene Zahl („Ich bin ____ Jahre“)

Zuordnungscode:

„Geben Sie jene Nummer ein, welche Ihnen vom Studienleiter zugewiesen wird.“

Antwortformat: offene Zahl („Ich bin Nummer ____“)

Knowledge-Skala (Wissen über KI/ChatGPT)

„Lesen Sie sich bitte jede Aussage aufmerksam durch und kreuzen Sie an, wie sehr Sie ihr zustimmen.“

Antwortformat (für alle Items):

5-stufige Likert-Skala, „stimme gar nicht zu“ bis „stimme voll zu“

Ausweichoption: „kann ich nicht beurteilen“

Items:

KH01_01: „ChatGPT kann mathematische Probleme zuverlässig lösen.“

KH01_02: „ChatGPT gibt auf dieselbe Frage immer dieselbe Antwort.“

KH01_03: „ChatGPT ist so programmiert, dass es wie ein Mensch schreibt.“

KH01_05: „ChatGPT schreibt manchmal Antworten, die nicht stimmen.“

KH01_06: „ChatGPT greift auf Informationen zu, die täglich aktualisiert werden.“

Use-Skala (Diversität der Nutzung)

„Lesen Sie sich bitte jede Aussage aufmerksam durch und kreuzen Sie an, wie häufig Sie ChatGPT in der Vergangenheit für die folgenden Tätigkeiten genutzt haben.“

Antwortformat (für alle Items):

5-stufige Likert-Skala, „nie“ bis „sehr häufig“, keine Ausweichoption

Items:

KH02_01: „Ich habe ChatGPT bereits genutzt, um Feedback für Hausaufgaben zu erhalten, die ich zu erledigen hatte.“

KH02_06: „Ich habe ChatGPT bereits genutzt, um Bilder analysieren zu lassen.“

KH02_02: „Ich habe ChatGPT bereits genutzt, um mir Inhalte erklären zu lassen, die mir im Unterricht nicht sofort klar waren.“

KH02_03: „Ich habe ChatGPT bereits genutzt, um Bilder zu erstellen.“

KH02_04: „Ich habe ChatGPT bereits genutzt, um Texte zusammenfassen zu lassen.“

KH02_05: „Ich habe ChatGPT bereits genutzt, um mir einen Überblick über ein Thema zu verschaffen, das mir neu war.“

KH02_08: „Ich habe ChatGPT bereits genutzt, um Ratschläge zu bekommen (z. B. zum Zeitmanagement für ein Schulprojekt).“

Use-behavior-Variable (Nutzungsverhalten)

„Bitte wählen Sie aus, wie häufig Sie ChatGPT verwenden.“

Antwortformat:

7-stufige Einfachauswahl, keine Ausweichoption

Antwortoptionen:

- ☐ „Nie“
- ☐ „Einmal im Monat“
- ☐ „Mehrere Male im Monat“
- ☐ „Einmal pro Woche“
- ☐ „Mehrere Male pro Woche“
- ☐ „Einmal am Tag“

- „Mehrere Male am Tag“

Attitude-Skala (Einstellungen gegenüber ChatGPT)

„Lesen Sie sich bitte jede Aussage aufmerksam durch und kreuzen Sie an, wie sehr Sie ihr zustimmen.“

Antwortformat (für alle Items):

5-stufige Likert-Skala, „stimme gar nicht zu“ bis „stimme voll zu“, keine Ausweichoption

Items:

KH03_01: „ChatGPT ist zuverlässig.“

KH03_02: „ChatGPT ist unzuverlässig beim Umgang mit persönlichen Daten.“

KH03_03: „ChatGPT kann Vorurteile oder einseitige Ansichten verbreiten.“

KH03_04: „Die Nutzung von ChatGPT führt zu Problemen mit Urheberrechten.“

KH03_05: „ChatGPT braucht sehr viel Strom und Energie.“

KH03_06: „Chatbots wie ChatGPT gehören zu den wichtigsten Erfindungen des 21. Jahrhunderts.“

KH03_07: „Man weiß nicht genau, aus welchen Quellen ChatGPT seine Antworten zieht.“

KH03_08: „ChatGPT kann falsche Informationen weiterverbreiten.“

Self-efficacy-Skala (Selbstwirksamkeit)

„Lesen Sie sich bitte jede Aussage aufmerksam durch und kreuzen Sie an, wie sehr Sie ihr zustimmen.“

Antwortformat (für alle Items):

5-stufige Likert-Skala, „stimme gar nicht zu“ bis „stimme voll zu“, keine Ausweichoption

Items:

SH01_01: „Ich kann ChatGPT verwenden, um in der Schule besser zu werden.“

SH01_02: „Ich kann mit Hilfe von ChatGPT und etwas Mühe neue Sachen lernen.“

SH01_03: „Ich habe Interesse daran, mehr über ChatGPT und ähnliche Technologien zu lernen.“

SH01_04: „ChatGPT im Unterricht zu verwenden ist für mich schon ganz normal geworden.“

SH01_05: „Ich nutze ChatGPT, um meine Mitschriften oder Hausübungen zu überarbeiten.“

SH01_06: „Ich finde, dass ChatGPT heutzutage weit verbreitet ist.“

SH01_07: „Obwohl es manchmal aufwendig ist, ChatGPT zu verwenden, finde ich überwiegen die Vorteile daran.“

Trust-Skala (Vertrauen in ChatGPT)

„Lesen Sie sich bitte jede Aussage aufmerksam durch und kreuzen Sie an, wie sehr Sie ihr zustimmen.“

Antwortformat (für alle Items):

5-stufige Likert-Skala, „stimme gar nicht zu“ bis „stimme voll zu“, keine Ausweichoption

Items:

SH02_06: „Ich glaube, dass es sicher genug ist, mit ChatGPT zu schreiben.“

SH02_01: „Ich vertraue darauf, dass alles, was ich bei ChatGPT mache, privat und geschützt bleibt.“

SH02_02: „Ich denke, dass ChatGPT meine persönlichen Daten vertraulich behandelt.“

SH02_03: „Ich bin überzeugt, dass ChatGPT meine Daten vor neugierigen Blicken schützt.“

SH02_04: „Ich hoffe, dass ChatGPT inzwischen eine sichere Umgebung bietet.“

SH02_05: „Ich glaube, dass ChatGPT mir beim Planen von Unterricht oder Lernen helfen kann.“