

Schneider, Swen (Ed.)

**Working Paper**

## Künstliche Intelligenz an Hochschulen: Empirische Perspektiven zu generativer KI bei Abschlussarbeiten und in der Lehre

Working Papers, No. 38

**Provided in Cooperation with:**

Faculty of Business and Law, Frankfurt University of Applied Sciences

*Suggested Citation:* Schneider, Swen (Ed.) (2025) : Künstliche Intelligenz an Hochschulen: Empirische Perspektiven zu generativer KI bei Abschlussarbeiten und in der Lehre, Working Papers, No. 38, <https://doi.org/10.48718/bcqw-c740>

This Version is available at:

<https://hdl.handle.net/10419/330669>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

# Künstliche Intelligenz an Hochschulen

## Empirische Perspektiven zu generativer KI bei Abschlussarbeiten und in der Lehre

Swen Schneider (Hrsg.)

Working Papers

Fachbereich Wirtschaft und Recht

Nr. 38

Frankfurt University of Applied Sciences

**Frankfurt University of Applied Sciences**  
**Fachbereich 3: Wirtschaft und Recht**  
**Nibelungenplatz 1**  
**60318 Frankfurt am Main**

Das Urheberrecht liegt bei den Autor\*innen.

Working Papers des Fachbereichs Wirtschaft und Recht der Frankfurt University of Applied Sciences dienen der Verbreitung von Forschungsergebnissen aus laufenden Arbeiten im Vorfeld einer späteren Publikation. Sie sollen den Ideenaustausch und die akademische Debatte befördern. Die Zugänglichmachung von Forschungsergebnissen in einem Fachbereichs Working Paper ist nicht gleichzusetzen mit deren endgültiger Veröffentlichung und steht der Publikation an anderem Ort und in anderer Form ausdrücklich nicht entgegen.

Working Papers, die vom Fachbereich Wirtschaft und Recht herausgegeben werden, geben die Ansichten des/der jeweiligen Autor\*innen wieder und nicht die der gesamten Institution des Fachbereichs Wirtschaft und Recht oder der Frankfurt University of Applied Sciences.

Autor\*innen:

Die Autor\*innen werden in den jeweiligen Artikeln genannt: Alia Anjum Rizvi, Angela Altomare, Anuja Ariyarasa, Chris Meier, Hamdan Razaq, Julia Gimbel, Kerem Tekin, Madinah Mehrabuddin, Marcel Crone, Paula Vanselow, Philip Kreis, Swen Schneider, Zakaria Salah

Bitte zitieren als:

Schneider, Swen (Hrsg) (2025): Künstliche Intelligenz an Hochschulen. Working Paper Nr. 38 des Fachbereichs 3 Wirtschaft und Recht. Frankfurt University of Applied Sciences.  
DOI: <https://doi.org/10.48718/bcqw-c740>

## Vorwort

Masterstudierende im Bereich Digital Business Management an der Frankfurt University of Applied Sciences führten ein einjähriges empirisches Forschungsprojekt durch, um die Auswirkungen generativer Künstlicher Intelligenz (GenKI) auf das Hochschulwesen und Studium zu untersuchen. Grundlage waren Befragungen unter Studierenden und Dozierenden sowie eine Analyse von Abschlussnoten.

Das Projekt untersucht die Wahrnehmung und Nutzung von GenKI durch Studierende und Lehrende, ihre Einbindung in Lehr- und Prüfungsprozesse sowie die Auswirkungen auf akademische Leistungen. Ziel ist es, Chancen und Herausforderungen des KI-Einsatzes im Hochschulbereich aufzuzeigen und Impulse für die zukünftige Integration in die akademische Praxis zu geben. Obwohl die Studien nicht repräsentativ sind, geben sie doch erste Hinweise darauf, wie generative KI insbesondere in die Lehre eingebunden werden kann.

Ein zentrales Ergebnis ist das sogenannte KI-Paradoxon: Je intensiver Studierende GenKI nutzen, desto stärker wächst ihr Vertrauen in die Technologie – selbst bei dem Bewusstsein, dass KI fehlerhafte oder irreführende Inhalte erzeugen kann. Mit zunehmender Erfahrung steigt der Einsatz in akademischen Arbeiten, insbesondere in Abschlussarbeiten, während gleichzeitig die Bereitschaft sinkt, die generierten Inhalte kritisch zu prüfen.

Swen Schneider



Vordere Reihe: Alia Anjum Rizvi, Madinah Mehrabuddin, Angela Altomare, Anuja Ariyarasa, Julia Gimbel, Paula Vanselow. Hintere Reihe: Swen Schneider, Hamdan Razaq, Chris Meier, Philip Kreis, Kerem Tekin, Marcel Crone, Zakaria Salah.

## **Artikelverzeichnis**

**Einsatz von KI in Lehre und Forschung ..... 14**

Swen Schneider

**Der Einfluss generativer KI auf Abschlussarbeiten und  
Kolloquien an der Frankfurt UAS aus Sicht der Dozenten ..... 42**

Marcel Crone, Philip Kreis

**Eigenleistung bei wissenschaftlichen Arbeiten im Umbruch..... 63**

Anuja Ariyarasa, Zakaria Salah

**KI in schriftlichen Prüfungsformaten ..... 84**

Madinah Mehrabuddin, Alia Anjum Rizvi, Hamdan Razaq

**Vertrauen in Generative KI.....101**

Julia Gimbel

**KI in der Prüfungsvorbereitung .....122**

Angela Altomare, Paula Vanselow

**KI-Tutorensysteme.....146**

Chris Meier, Kerem Tekin

# Inhaltsverzeichnis

<b>1</b>	<b>Einsatz von KI in Lehre und Forschung .....</b>	<b>14</b>
1.1	KI in der Lehre .....	14
1.1.1	Automated Content Generation mit Large Language Models (LLMs) .....	14
1.1.2	Automated Decision Making und Automated Essay Scoring (AES).....	16
1.1.3	Mit Avataren als Ergänzung von Dozenten zum Omnichannel-Learning .....	18
1.1.4	Neue Kriterien für die Bewertung von Abschlussarbeiten.....	20
1.2	KI in der Forschung.....	22
1.2.1	Wissensproduktion im Zeitalter Künstlicher Intelligenz .....	22
1.2.2	Transformation in der Forschung.....	23
1.3	Die Zukunft des Lehrens und Lernens an Hochschulen .....	24
1.3.1	Kommt es zu einer Hyperautomatisierung in der Hochschule? .....	24
1.3.2	Entwicklung von Future Skills im Hochschulkontext.....	25
1.4	Fazit: Was muss man zukünftig noch Wissen, wenn ChatGPT jederzeit und überall verfügbar ist? .....	26
1.5	Literaturverzeichnis .....	27
<b>2</b>	<b>Grundlagen als Ausgangspunkt empirischer Untersuchungen.....</b>	<b>29</b>
2.1	Begriffe im Zeitalter generativer KI.....	29
2.1.1	Künstliche Intelligenz .....	29
2.1.2	Definition generative KI und Einordnung im wissenschaftlichen Arbeiten.....	29
2.1.3	Institutionelle Regelungen und rechtlicher Rahmen.....	30
2.1.4	ChatGPT .....	31
2.2	Datenerhebung und Stichprobe.....	31
2.2.1	Forschungsdesign und Vorgehen .....	31
2.2.2	Forschungstücke .....	32
2.2.3	Erhebungsinstrument und Datenerhebung .....	32
2.3	Deskriptive Statistiken .....	33
2.3.1	Soziodemografische Statistiken .....	33
2.3.2	Nutzung und Erfahrung mit generativer KI im Rahmen des Studiums.....	34
2.3.3	Einfluss von KI .....	37
2.4	Limitationen der Studien.....	37
2.4.1	Limitation der Arbeit: „GenKI bei Abschlussarbeiten aus Sicht der Dozenten“ .....	37
2.4.2	Limitationen der Arbeit „GenKI bei Abschlussarbeiten aus Sicht der Studierenden“ .....	38
2.5	Literaturverzeichnis .....	39

<b>Generative KI bei Abschlussarbeiten aus Sicht der Dozenten .....</b>	<b>41</b>
<b>3 Der Einfluss generativer KI auf Abschlussarbeiten und Kolloquien an der Frankfurt UAS aus Sicht der Dozenten .....</b>	<b>42</b>
3.1 Noten unter Beobachtung: Der Einfluss von KI auf die Hochschulprüfung.....	43
3.1.1 Zielsetzung der Arbeit.....	43
3.1.2 Aufbau der Arbeit.....	44
3.2 Methodisches Vorgehen zur Analyse von KI-Einflüssen auf Prüfungsprozesse.....	44
3.2.1 Methodik zur Analyse der Umfrage von Lehrenden der Frankfurt UAS.....	44
3.2.2 Methodik zur Analyse von Abschlussnoten.....	45
3.3 Was Lehrende wirklich denken: KI in Abschlussarbeiten – empirische Einblicke aus Umfrage und Datenanalyse .....	45
3.3.1 Auswertung der Umfrage von Lehrenden der Frankfurt UAS.....	46
3.4 Zwischen Wandel und Widerstand – Was die Ergebnisse zeigen .....	58
3.4.1 Diskussion zu den Ergebnissen der Umfrage von Lehrenden der Frankfurt UAS .....	59
3.4.2 Diskussion zu den Ergebnissen der historischen Daten.....	59
3.5 Zentrale Erkenntnisse und Implikationen.....	60
3.6 Literaturverzeichnis.....	61
<b>Generative KI bei Abschlussarbeiten aus Sicht der Studierenden.....</b>	<b>62</b>
<b>4 Eigenleistung bei wissenschaftlichen Arbeiten im Umbruch.....</b>	<b>63</b>
4.1 Zwischen Anspruch und Anpassung .....	64
4.2 Grundlagen zu Eigenleistung und KI im wissenschaftlichen Arbeiten .....	65
4.2.1 Definition Eigenleistung.....	65
4.2.2 Psychologische und soziale Einflussfaktoren der KI in der Hochschullehre ..	66
4.2.3 Eigenleistung und KI-Nutzung im Fokus aktueller empirischer Forschung.....	67
4.3 Methodisches Vorgehen .....	68
4.3.1 Datenanalyse .....	68
4.3.2 Kodierung und Datenaufbereitung.....	69
4.3.3 Statistische Verfahren .....	70
4.4 Eigenleistung unter dem Mikroskop .....	70
4.4.1 Zusammenhang zwischen KI-Nutzung und kritischer Reflexion.....	71
4.4.2 Zusammenhang zwischen Selbstwirksamkeit und Eigenleistungsverständnis ..	71
4.4.3 Zusammenhang zwischen Regelwissen, sozialem Rückhalt und KI-Abgrenzung.....	73
4.4.4 Weitere Ergebnisse .....	75
4.4.5 Geeignete Nachweismethoden.....	78
4.5 Eigenleistung im Spiegel der Reflexion .....	79
4.5.1 Kritische Reflexion von Eigenleistung im Umgang mit KI .....	79
4.5.2 Selbstwirksamkeit und Haltung zu Eigenleistung .....	79

4.5.3	Institutionelle Regelkenntnis und sozialer Einfluss.....	80
4.5.4	Präferenzen bei Nachweismethoden und hochschuldidaktische Implikationen .....	80
4.5.5	Methodenkritik und Kontextualisierung.....	81
4.6	Neuverhandlung wissenschaftlicher Eigenleistung im KI-Zeitalter .....	81
4.7	Literaturverzeichnis.....	82
<b>5</b>	<b>KI in schriftlichen Prüfungsformaten.....</b>	<b>84</b>
5.1	Prüfungskultur im Wandel: Zur Notwendigkeit dieser Untersuchung.....	85
5.1.1	Relevanz des Themas.....	85
5.1.2	Zielsetzung und Forschungsfragen .....	85
5.1.3	Struktur der Untersuchung .....	86
5.2	KI in Hochschulen im Spiegel von Praxis und Forschung.....	86
5.2.1	Institutionelle Regelungen und rechtlicher Rahmen .....	86
5.2.2	Aktueller Forschungsstand .....	87
5.3	Von der Theorie zur Datengrundlage .....	88
5.4	Empirische Einblicke in studentische Perspektiven .....	88
5.4.1	Deskriptive Ergebnisse der Befragung.....	89
5.4.2	Vertiefende statistische Analysen.....	90
5.5	Zwischen Integration und Abgrenzung.....	96
5.6	Schlussfolgerungen und Implikationen für die Hochschulpraxis.....	98
5.7	Literaturverzeichnis.....	98
<b>6</b>	<b>Vertrauen in Generative KI.....</b>	<b>101</b>
6.1	Einleitung .....	102
6.2	Definitionen und Theorien .....	102
6.2.1	Validität einer Information .....	102
6.2.2	Vertrauen bei der Nutzung von ChatGPT .....	103
6.2.3	Schwerpunkt: Validität von Informationen.....	103
6.2.4	Überblick über bestehende Forschung zu Generativer KI in der Bildung.....	104
6.2.5	Vertrauensfragen in der Nutzung von KI .....	105
6.2.6	Vergleich mit anderen Technologien.....	105
6.2.7	Abgeleitete Hypothesen .....	106
6.3	Methodik .....	106
6.3.1	Forschungsdesign und -ansatz .....	106
6.3.2	Datenerhebung.....	107
6.3.3	Datenanalyseverfahren .....	107
6.4	Auswertung .....	108
6.4.1	Deskriptive Statistiken.....	108
6.4.2	Korrelationen.....	111
6.4.3	Hypothesentests .....	112
6.4.4	Multiple Regressionsanalyse .....	115

6.5	Diskussion.....	119
6.6	Fazit.....	120
6.7	Literaturverzeichnis.....	120
<b>7</b>	<b>KI in der Prüfungsvorbereitung .....</b>	<b>122</b>
7.1	Themenhintergrund und Zielsetzung .....	123
7.1.1	Themenrelevanz .....	123
7.1.2	Zielsetzung der Arbeit und Problemstellung .....	124
7.1.3	Forschungsfrage und Hypothesen.....	124
7.1.4	Aufbau des Artikels .....	124
7.2	Grundlagen der Prüfungsangst und Veränderungen der Kompetenzen: Theoretische Modelle und Forschungslücken.....	125
7.2.1	Prüfungsangst: Definitionen und Einflussfaktoren .....	125
7.2.2	Veränderungen der Kompetenzen durch generative KI in der Hochschulbildung .....	125
7.2.3	Forschungsstand und Forschungslücke .....	126
7.3	Auf dem Weg zur Erkenntnis.....	127
7.3.1	Forschungslücke .....	127
7.3.2	Erhebungsinstrumente und Operationalisierung .....	127
7.3.3	Datenerhebung und statistische Auswertungsverfahren .....	128
7.3.4	Gütekriterien .....	128
7.4	Ergebnisse der Datenanalyse .....	129
7.4.1	Deskriptive Statistiken.....	129
7.4.2	Gruppenvergleiche .....	135
7.4.3	Zentrale Befunde .....	139
7.5	Zwischen Erkenntnis und Bedeutung .....	141
7.5.1	Interpretation der Ergebnisse im Kontext der Forschungsfrage .....	141
7.5.2	Einordnung in den theoretischen und empirischen Forschungsstand.....	142
7.5.3	Praktische Implikationen und Forschungsperspektiven.....	142
7.6	Vom Ausgangspunkt zu den zentralen Tendenzen.....	143
7.6.1	Zusammenfassung der wichtigsten Erkenntnisse .....	143
7.6.2	Beantwortung der Forschungsfrage und Hypothesen.....	143
7.7	Literaturverzeichnis.....	144
<b>8</b>	<b>KI-Tutorensysteme .....</b>	<b>146</b>
8.1	Zwischen Akzeptanz und Ablehnung – Hochschulen im KI-Dilemma .....	147
8.1.1	Lernen neu gedacht – Wie KI den Bildungssektor verändert .....	147
8.1.2	Studieren mit System – KI-Tutoren unter der Lupe.....	148
8.1.3	Ersetzt KI den Dozenten? – Die zentrale Frage dieser Arbeit.....	148
8.2	Wissen schafft Verständnis – Die Theorie hinter dem KI-Tutor .....	148
8.2.1	Was genau ist ein KI-Tutor? Begriffe und Konzepte im Überblick .....	148
8.2.2	Studienlage zu KI-Tutoren im Überblick .....	150

8.3	KI statt Präsenz?.....	152
8.3.1	Forschungsfragen .....	152
8.3.2	Hypothesen.....	152
8.3.3	Erwartete Effekte .....	153
8.4	Der Weg zur Antwort.....	153
8.4.1	Forschungsdesign.....	153
8.4.2	Stichprobe .....	154
8.4.3	Erhebungsinstrumente.....	154
8.5	Nutzung ja – Zahlung nein .....	155
8.5.1	Datenanalyse .....	155
8.5.2	Statistische Tests.....	158
8.5.3	Ergebnisinterpretation.....	160
8.6	KI überzeugt – mit Grenzen.....	161
8.6.1	Kernerkenntnisse .....	161
8.6.2	Handlungsempfehlungen .....	162
8.6.3	Ausblick .....	162
8.7	Literaturverzeichnis.....	162
<b>9</b>	<b>Kernfazit dieser Veröffentlichung.....</b>	<b>164</b>
<b>10</b>	<b>Bisher erschienene Working Papers des Fachbereich 3 .....</b>	<b>165</b>

## Abbildungsverzeichnis

Abbildung 1.1:	Der Turing Test .....	15
Abbildung 1.2:	LLM (Large Language Modells) und ihre Funktionsweise .....	16
Abbildung 1.3:	Automatisierte Entscheidungen ( <a href="https://dataethics.eu/cpdp-2017-ethics-age-intelligent-machines/">https://dataethics.eu/cpdp-2017-ethics-age-intelligent-machines/</a> ) .....	17
Abbildung 1.4:	Omnichannel-Learning .....	19
Abbildung 1.5:	Bewertung von Abschlussarbeiten .....	21
Abbildung 1.6:	Quellenvisualisierung ( <a href="https://researchrabbitapp.com">https://researchrabbitapp.com</a> ) 08/2025 .....	22
Abbildung 1.7:	Kriterien zum Einsatz von KI an Hochschulen (Chan und Colloton 2024) .....	25
Abbildung 1.8:	Menschliche Interaktion mit der KI.....	27
Abbildung 2.1:	Altersverteilung der Umfrageteilnehmende .....	33
Abbildung 2.2:	Geschlechterverteilung der Umfrageteilnehmende.....	34
Abbildung 2.3:	Semesterverteilung der Umfrageteilnehmende.....	34
Abbildung 2.4:	Anzahl an geschriebenen Hausarbeiten pro Jahr in % .....	35
Abbildung 2.5:	Einschätzung der Erfahrung der Studierenden im Umgang mit ChatGPT .....	36
Abbildung 3.1:	Zusammenhang von Veränderungen und Einfluss von KI .....	43
Abbildung 3.2:	Wahrgenommene Veränderung seit der Verfügbarkeit von KI .....	46
Abbildung 3.3:	Veränderung der intellektuellen Leistung, um eine Abschlussarbeit zu schreiben.....	47
Abbildung 3.4:	Vertrauen auf die Eigenständigkeit studentischer Abschlussarbeiten .....	47
Abbildung 3.5:	Veränderung der Qualität der Quellen in Abschlussarbeiten seit der Verfügbarkeit von generativer KI .....	48
Abbildung 3.6:	Wahrnehmung von fiktiven oder falschen Quellen seit der Verfügbarkeit von generativer KI .....	48
Abbildung 3.7:	Veränderung der Bewertungskriterien bei Abschlussarbeiten seit der Verfügbarkeit generativer KI.....	49
Abbildung 3.8:	Ermutigung von Studierenden zum Einsatz von generativer KI in der Bearbeitung von Abschlussarbeiten .....	51
Abbildung 3.9:	Bedenken bezüglich der Ermutigung zur Nutzung generativer KI in Abschlussarbeiten .....	52
Abbildung 3.10:	Anzahl der Datenpunkte pro Prüfungssemester und Prüfungsart (Vor Dateneinschränkung) .....	54
Abbildung 3.11:	Anzahl der Datenpunkte pro Prüfungssemester und Prüfungsart (Vor Dateneinschränkung) .....	54
Abbildung 3.12:	Notenentwicklung mit gleitendem 4-Semester Durchschnitt.....	55
Abbildung 3.13:	Verteilung der Dateneinträge nach Kategorie .....	55
Abbildung 3.14:	Kumulative Verteilung der Thesis-Noten.....	56
Abbildung 3.15:	Kumulative Verteilung der Kolloquium-Noten .....	58
Abbildung 3.16:	Umfrageergebnis - Wissensdiskrepanz zwischen Kolloquium und Abschlussarbeit.....	59
Abbildung 4.1:	Zusammenhang zwischen Selbstwirksamkeit und Eigenleistungsverständnis..	73
Abbildung 4.2:	Zusammenhang zwischen Regelkenntnis und Offenlegungsbereitschaft .....	74
Abbildung 4.3:	Zusammenhang zwischen sozialem Einfluss und Offenlegungsbereitschaft .....	75
Abbildung 4.4:	Zusammenhang zwischen Selbstwirksamkeit und Eigenleistungsverständnis .....	76
Abbildung 4.5:	Unterschiede nach Geschlecht in Bezug auf Offenlegung und Eigenleistung .....	77

Abbildung 4.6:	Zusammenhang zwischen der Erfahrung mit Nutzung von ChatGPT und des Eigenleistungsverständnisses .....	77
Abbildung 4.7:	Zusammenhang zwischen der Erfahrung mit Nutzung von ChatGPT und der Offenlegungsbereitschaft .....	78
Abbildung 5.1:	Neugewichtung von Aspekten in Hausarbeiten.....	85
Abbildung 5.2:	Visuelle Darstellung der Cluster .....	96
Abbildung 6.1:	Ängste und Bedenken zur Nutzung von ChatGPT im Studium .....	101
Abbildung 6.2:	Komponentenübersicht.....	104
Abbildung 6.3:	Beurteilung der Vertrauenswürdigkeit von ChatGPT-Informationen für akademische Arbeiten .....	108
Abbildung 6.4:	Bedeutung der Validität von ChatGPT generierten Informationen für akademische Arbeiten .....	109
Abbildung 6.5:	Zeitaufwand, der Studierenden zur Validierung der Ergebnisse von ChatGPT .....	109
Abbildung 6.6:	Ängste und Bedenken hinsichtlich der Nutzung von ChatGPT im Studium .....	110
Abbildung 6.7:	Kurzbericht: Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten .....	111
Abbildung 7.1	Zusammenhang zwischen Prüfungsangst und Nutzungshäufigkeit von ChatGPT .....	123
Abbildung 7.2:	Absolute Häufigkeit der Nutzungsalternativen von ChatGPT bei Klausurvorbereitung.....	129
Abbildung 7.3:	Häufigkeit der Nutzung von klassischen Lernmethoden ohne ChatGPT .....	130
Abbildung 7.4:	Häufigkeitsverteilungen der Bewertungen wichtiger Fähigkeiten für die Nutzung von ChatGPT bei der Klausurvorbereitung.....	131
Abbildung 7.4:	Häufigkeitsverteilung Aussagen zur Prüfungsangst.....	133
Abbildung 7.6:	Häufigkeitsverteilung - Aussagen zu ChatGPT im Prüfungskontext.....	134
Abbildung 8.1:	Zusammenhänge zur Einschätzung von KI-Tutorensystemen .....	147
Abbildung 8.2:	Häufigkeitsverteilung Semester.....	156
Abbildung 8.3:	Häufigkeitsverteilung der Antworten zur KI-Nutzungshäufigkeit.....	156
Abbildung 8.4:	Häufigkeitsverteilung der Antworten zu KI-Erfahrung .....	157
Abbildung 8.4	Spearman-Korrelationsanalyse der Variablen im Kontext KI-gestützte Tutorensysteme .....	158
Abbildung 8.6	Häufigkeitsverteilung der Antworten zur Variable KI-Nutzung .....	159
Abbildung 8.7	Zusammengefasste Darstellung des Hypothesentests auf Basis der Spearman-Korrelationsanalyse .....	159

## **Tabellenverzeichnis**

Tabelle 2.1:	Bewertung der eigenen ChatGPT-Erfahrung durch Studierende (N=66).....	35
Tabelle 2.2:	Häufigkeit der Nutzung von ChatGPT im universitären Kontext pro Jahr (N=66).....	36
Tabelle 2.3:	Einschätzung der Auswirkungen von KI auf Hausarbeiten .....	37
Tabelle 3.1:	Kreuztabelle Bewertungsmaßnahmen .....	50
Tabelle 3.2	Kreuztabelle Alltägliche Nutzung von KI und Ermutigung von KI-Einsatz durch Studierende.....	51
Tabelle 3.3:	Kreuztabelle Ermutigung zur Nutzung von KI und Selbsteinschätzung des Wissens über generative KI .....	52
Tabelle 3.4:	Ergebnisse WLS-Regression für Hypothese 4 .....	57
Tabelle 3.5:	Ergebnisse WLS-Regression für Hypothese 5 .....	58
Tabelle 4.1:	Bewertung der Eigenleistung bei Nutzung von KI.....	63
Tabelle 4.2:	Kodierung soziodemografische Variablen.....	69
Tabelle 4.3:	Kodierung themenspezifische Kontextvariablen.....	70
Tabelle 4.4:	Korrelation zwischen Nutzung_ChatGPT und Reflexion_Kritik .....	71
Tabelle 4.4:	Nachweismethoden.....	79
Tabelle 5.1:	Übersicht der analysierten Hochschulen .....	87
Tabelle 5.2:	Einschätzung der Auswirkungen von KI auf Hausarbeiten .....	89
Tabelle 5.3:	Bewertung von Anforderungen im KI-Kontext .....	89
Tabelle 5.4:	Zustimmung zu hochschulischen Maßnahmen .....	90
Tabelle 5.5:	Bewertung alternativer Prüfungsformate .....	90
Tabelle 5.6:	Übersicht relevanter Korrelationen ( $ r  \geq 0,30$ ) .....	91
Tabelle 5.7:	Gruppenunterschiede nach Geschlecht (Kruskal-Wallis, Mann-Whitney-U).....	92
Tabelle 5.8:	Gruppenunterschiede nach Erfahrung (Kruskal-Wallis-Test) .....	93
Tabelle 5.9:	Gruppenunterschiede nach Haltung (Kruskal-Wallis-Test) .....	93
Tabelle 5.10:	Gruppenunterschiede nach Maßnahmen (Kruskal-Wallis-Test) .....	94
Tabelle 5.11:	Ergebnisse der Regressionsanalyse .....	94
Tabelle 5.12	Ergebnis der Clusteranalyse .....	95
Tabelle 6.1:	Pearson Korrelationsmatrix .....	112
Tabelle 6.2:	Ergebnisse des Kruskal-Wallis-Tests zu den Hypothesen 1 und 2 .....	113
Tabelle 6.3:	Teststatistiken des Kruskal-Wallis Tests zu den Hypothesen 1 und 2.....	113
Tabelle 6.4:	Ergebnis des Kruskal-Wallis-Tests zur Hypothese 3 .....	114
Tabelle 6.5:	Teststatistiken des Kruskal-Wallis Tests zur Hypothese 3 .....	114
Tabelle 6.5:	Ergebnis des Kruskal-Wallis-Tests zur Hypothese 4 .....	115
Tabelle 6.7:	Teststatistiken des Kruskal-Wallis Tests zur Hypothese 4 .....	115
Tabelle 6.8:	Variablenkürzel der Regressionsgleichungen .....	116
Tabelle 6.9:	Regressionsgleichungen.....	116
Tabelle 6.10:	Koeffizienten der Regressionsgleichungen .....	117
Tabelle 6.11:	Model Zusammenfassung der Regression .....	117
Tabelle 6.12:	Analyse der Varianz der Regressionsgleichungen.....	118
Tabelle 7.1:	Deskriptive Statistiken zur Nutzungsalternativen von ChatGPT bei Klausurvorbereitung (Quelle: JASP).....	130
Tabelle 7.2:	Deskriptive Statistiken für klassische Lernmethoden bei der Klausurvorbereitung ohne ChatGPT (Quelle: JASP) .....	131
Tabelle 7.3:	Deskriptive Statistiken zur Bewertung relevanter Fähigkeiten für die Nutzung von ChatGPT (Quelle: JASP) .....	132
Tabelle 7.4:	Deskriptive Statistiken Aussagen zu unterschiedlichen Dimensionen der Prüfungsangst (Quelle: JASP) .....	133

Tabelle 7.5:	Deskriptive Statistiken Gesamt Prüfungsangst-Score (Quelle: JASP).....	134
Tabelle 7.6:	Deskriptive Statistiken - Aussagen zu ChatGPT im Prüfungskontext (Quelle: JASP) .....	135
Tabelle 7.7:	Gruppenvergleiche zwischen männlichen und weiblichen Teilnehmenden hinsichtlich Prüfungsangst.....	136
Tabelle 7.8:	Gruppenvergleiche zwischen männlichen und weiblichen Teilnehmenden ChatGPT-bezogenen Aussagen .....	137
Tabelle 7.9:	Kruskal-Wallis-Test: Vergleich von Semester und Aussagen zur Prüfungsangst (Quelle:JASP) .....	138
Tabelle 7.10:	Kruskal-Wallis-Test Vergleich von Semester und Aussagen zu ChatGPT im Prüfungskontext (Quelle:JASP) .....	138
Tabelle 7.11:	Dunns-Post-hoc-Test von Semester und Aussagen zu ChatGPT im Prüfungskontext (Quelle: JASP) .....	139
Tabelle 7.12:	Spearman Korrelation zwischen Gesamt-Prüfungsangstscore und Häufigkeit der Nutzung von ChatGPT im universitären Kontext (Quelle: JASP) .....	139
Tabelle 7.13:	Spearman Korrelation zwischen Prüfungsangst-Aussagen und Einschätzung zur ChatGPT-Nutzung im Prüfungskontext (Quelle: JASP).....	140
Tabelle 8.1:	Operationalisierung Hypothesen und Fragen und Zielvariablen.....	155

# 1 Einsatz von KI in Lehre und Forschung

Swen Schneider

## Abstract

Die rasante Entwicklung generativer KI verändert Lehre und Forschung an Hochschulen grundlegend. Automatisierte Inhalte, Bewertungen und interaktive Avatare erweitern Lern- und Prüfungskulturen, während neue Kompetenzen wie Prompting, kritisches Denken und Data Literacy an Bedeutung gewinnen. **Die Bewertung von Abschlussarbeiten verlagern sich von reiner Inhaltsbewertung hin zu reflektierenden und prozessbezogenen Anteilen**, wodurch Bewertungsmaßstäbe neu definiert werden müssen. Diese Transformation markiert einen epistemologischen und didaktischen Paradigmenwechsel, der Hochschulen zu klaren Leitlinien und ethischen Standards im Umgang mit KI verpflichtet.

**Schlagwörter:** Prompting, Abschlussarbeiten, Automated Essay Scoring, Omni-Channel-Learning, Agentic AI, Hyperautomatisierung,

„A fool with a tool is still a fool“ – gilt diese Aussage auch noch im Zeitalter der Künstlichen Intelligenz? Mit der rasanten Entwicklung greift die Künstliche Intelligenz (KI) immer stärker in unseren Alltag ein und verändert nicht nur unser Verhalten, sondern auch unsere Gewohnheiten.

Generative KI-Systeme wie ChatGPT, GoogleGemini oder Grok, etc. nehmen dabei auch im akademischen Bereich zunehmend Raum ein. Im Folgenden wird untersucht, wie

1. generative KI die Lehre und das Lernen sowie
2. die Erstellung und Korrektur von Abschlussarbeiten

beeinflusst.

Als Ergebnis kann festgehalten werden, dass durch die zunehmende **Hyper-Automatisierung in Lehre und Forschung** traditionelle Fähigkeiten (wie Social Skills oder kritisches Denken) als auch technisches Anwenderwissen (wie Prompting und Tooling) an Bedeutung gewinnen werden. Der Schwerpunkt von Abschlussarbeiten verlagert sich zunehmend **von inhaltlicher Ausarbeitung hin zu reflektierenden und prozessbezogenen Anteilen**. Die Bewertungskriterien werden daher vermehrt den Weg der Erkenntnisgenerierung als das fertige Ergebnis in den Mittelpunkt stellen.

## 1.1 KI in der Lehre

### 1.1.1 Automated Content Generation mit Large Language Models (LLMs)

Computer „denken“ nicht im gleichen Sinne wie Menschen. Als Alan Turing 1950 den **Turing-Test** vorschlug, definierte er künstliche Intelligenz so: Eine Maschine gilt als intelligent, wenn ein Mensch A in einer Unterhaltung nicht mehr zwischen den Antworten der Maschine und denen eines anderen Menschen B unterscheiden kann (siehe Bild unten). Obwohl dieser Test über 70 Jahre lang als Maßstab galt, haben bahnbrechende Innovationen und rasante Fortschritte in der Künstlichen Intelligenz (KI) diese Perspektive erst jetzt verändert. Moderne KI-Systeme verfügen über Fähigkeiten, die weit über das hinausgehen, was Turing damals für möglich hielt, und werfen die Frage auf, wie wir „Denken“ im Zusammenhang mit Computern heute neu definieren sollten (Turing, 1950).

Der Turing Test als Prüfstein der KI ist nun in Zeiten der LLMs zur Realität geworden

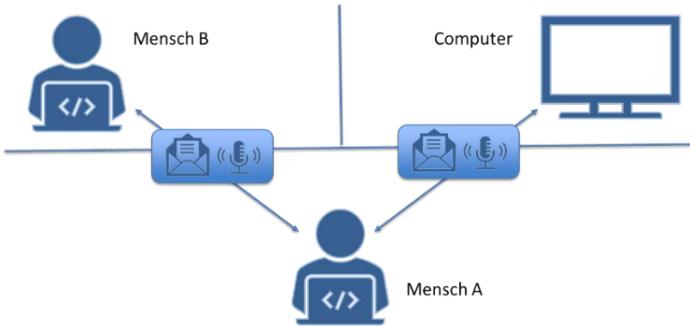


Abbildung 1.1: Der Turing Test

Während allgemeine Künstliche Intelligenz ein breites Fähigkeitsspektrum abdeckt, fokussiert sich **generative KI** speziell auf Sprache, Text und Bild, indem sie Wahrscheinlichkeiten für Wörter berechnet und vor allem das nächste Token vorhersagt. Dieser Prozess beginnt mit der **Tokenisierung**, bei der der Eingangstext in kleinere Einheiten zerlegt wird, die sogenannten Tokens. Diese Tokens können dabei Wörter, Wortteile oder auch Symbole sein, je nachdem, wie die Tokenisierung angelegt ist. Anschließend werden die Tokens in Embeddings umgewandelt, die numerische Repräsentationen der Tokens in einem hochdimensionalen Raum darstellen. Diese **Embeddings** erlauben es dem Modell, die Bedeutung der Tokens zu erfassen und ihre kontextuellen Beziehungen zu verstehen. Die Embeddings werden dann durch ein neuronales Netz, in der Regel ein Transformer-Modell, verarbeitet. Diese Modelle bestehen aus mehreren Schichten, die Transformationen auf die Daten anwenden und dabei lernen, wie bestimmte Tokens in verschiedenen Kontexten miteinander interagieren. Ein prominentes Beispiel für solch ein Modell ist GPT (**Generative Pre-trained Transformer**). Wie in der nachstehenden Grafik dargestellt sind GPT-Modelle darauf trainiert, riesige Textmengen zu analysieren und zu verstehen, welche Wörter und Phrasen in bestimmten Kontexten häufig auftreten, um dann den nächsten Token, basierend auf der bisherigen Eingabe, vorherzusagen (Buck 2025).

Ein wichtiger Bestandteil dieses Prozesses ist das **Prompting**, bei dem der Nutzer dem Modell eine Eingabeaufforderung gibt, die als Ausgangspunkt für die Textgenerierung dient. Durch präzises Prompting kann das Modell gezielt auf spezifische Anforderungen reagieren und passende Texte generieren. Auf diese Weise wird der Text Schritt für Schritt aufgebaut, indem das Modell für jeden neuen Schritt die Wahrscheinlichkeit für den nächsten Token berechnet und den wahrscheinlichsten Token auswählt (Paaß 2023).

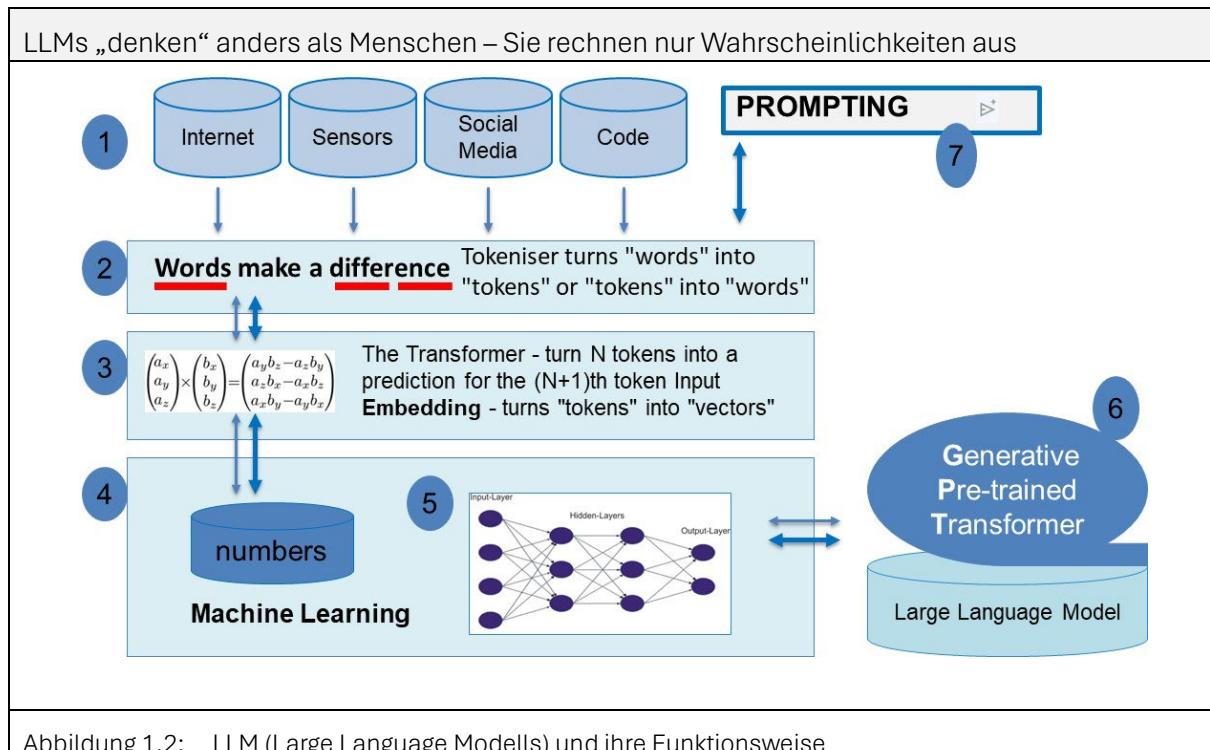


Abbildung 1.2: LLM (Large Language Models) und ihre Funktionsweise

Mit der heutigen Hardware, insbesondere den leistungsstarken Grafikchips, ist die Entwicklung von Large Language Models wie ChatGPT oder Google Gemini nun Realität geworden. Diese Technologien führen zunehmend dazu, dass wir Menschen immer weniger zwischen menschlich und künstlich erzeugten Inhalten unterscheiden können – etwa bei Fake News, gesprochener Sprache, Bildern oder künftig auch Videos (Deepfakes). In Zukunft wird es daher notwendig sein, **Intermediäre** einzusetzen, wie zum Beispiel **vertrauenswürdige KI-Agenten (Agentic AI)**, die vorab die Quellen und die Vertrauenswürdigkeit von Inhalten prüfen. Erst wenn diese Agenten die Informationen validiert haben und mit einer bestimmten Wahrscheinlichkeit als vertrauenswürdig einstufen, werden wir uns mit den Inhalten auseinandersetzen.

Auch in der Wissenschaft wird es für uns Menschen zunehmend schwieriger, von Personen verfasste Texte von maschinell generierten Inhalten zu unterscheiden. Daher wird es immer wichtiger, die Qualität und den Wahrheitsgehalt von Antworten kritisch zu überprüfen – insbesondere dann, wenn diese lediglich mit einer gewissen prozentualen Wahrscheinlichkeit richtig sein könnten.

### 1.1.2 Automated Decision Making und Automated Essay Scoring (AES)

Nicht nur die Generierung, sondern auch die Überprüfung und Bewertung von Texten kann zunehmend automatisiert erfolgen. Ein zentrales Anwendungsfeld stellt dabei das **Automated Essay Scoring (AES)** dar – ein Teilbereich der automatisierten Entscheidungsfindung (Araujo 2014), der sich auf die Bewertung schriftlicher Arbeiten, insbesondere im wissenschaftlichen und bildungsbezogenen Kontext, konzentriert. AES-Systeme nutzen Verfahren der natürlichen Sprachverarbeitung (NLP), des maschinellen Lernens (ML) und der künstlichen Intelligenz (KI), um Texte automatisiert zu analysieren und zu bewerten. Ziel ist es, menschliche Prüfer zu entlasten und gleichzeitig schnelles, konsistentes Feedback zu ermöglichen. Hierzu werden aus den eingereichten Texten Merkmalsvariablen extrahiert – etwa zur Grammatik, Rechtschreibung, Wortwahl, inhaltlichen Struktur oder stilistischen Ausprägung – und statistisch zu einer Gesamtnote

zusammengeführt (Ramesh 2022). Während frühere Systeme noch manuell definierte Regelwerke erforderten, ermöglicht der Einsatz generativer KI heute eine deutlich flexiblere, effizientere und benutzerfreundlichere Implementierung. Zahlreiche kommerzielle Anbieter bieten mittlerweile entsprechende Lösungen an (Ifenthaler 2025).

Der **europäische AI Act** schreibt vor, dass bei automatisierten Entscheidungsprozessen der Mensch die letztendliche Entscheidungsverantwortung tragen muss (Human-In-The-Loop). Dem steht jedoch der Ansatz der **algorithmischen Autorität** gegenüber. Dieser beschreibt die zunehmende Tendenz, algorithmischen Systemen aufgrund ihrer vermeintlichen Objektivität und Effizienz Entscheidungskompetenz und Legitimität zuzuschreiben. Dadurch werden menschliche Urteile zunehmend durch datengestützte, jedoch häufig intransparente Prozesse ersetzt, was die Nachvollziehbarkeit und kritische Auseinandersetzung erschwert. Da viele Menschen die zugrunde liegenden Daten und Algorithmen nicht vollständig verstehen und dies oft nicht eingestehen, folgen sie algorithmischen Empfehlungen meist ohne Widerspruch – selbst dann, wenn berechtigte Zweifel bestehen.

Bei automatisierten Entscheidungen soll es nach dem EU AI Act immer einen HITL (Human-In-The Loop), einen Menschen als letzte Entscheidungsinstanz geben

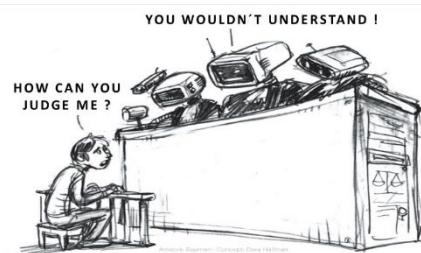


Abbildung 1.3: Automatisierte Entscheidungen (<https://dataethics.eu/cpdp-2017-ethics-age-intelligent-machines/>)

Automatisierte Entscheidungen finden dann Akzeptanz, wenn sie nachvollziehbar sind. Eine wichtige Forschungsrichtung in der KI ist daher die Erklärbare Künstliche Intelligenz (XAI – Explainable AI), die darauf abzielt, die Funktionsweise und Entscheidungen von KI-Systemen transparent zu machen. Selbst wenn eine Erklärungskomponente vorhanden ist, ist es entscheidend, dass eine menschliche Eskalationsstufe für die Klärung von Unstimmigkeiten besteht. Dies schafft Vertrauen und ermöglicht Korrekturen bei Fehlern oder unerwarteten Ergebnissen. Ein oft vorgebrachtes Argument für den Einsatz von KI ist, dass automatisierte Systeme weniger voreingenommen reagieren und somit neutralere Entscheidungen treffen können als Menschen (Haried und Schneider 2025). Dies könnte dazu beitragen, bestehende menschliche "Biases" (Voreingenommenheit) zu reduzieren. Es gibt aber auch das gegenteilige Argument, dass KI-Systeme Verzerrungen (Bias) enthalten können, weil sie auf historischen Daten basieren, die Vorurteile oder Ungleichheiten widerspiegeln. Dies kann zu unfairen oder diskriminierenden Ergebnissen führen und stellt ein häufig genanntes Risiko beim Einsatz von KI dar.

Transparenz, Nachvollziehbarkeit sowie die Möglichkeit zu Einspruch und menschlicher Klärung sind entscheidende Faktoren für die Akzeptanz automatisierter Entscheidungen. Automatisierte Essaybewertung (Automated Essay Scoring, AES) ist ein vergleichsweise junges Forschungs- und Anwendungsfeld. Entsprechend variieren die technologischen Ansätze und Schwerpunktsetzungen der Anbieter. Zu den führenden Unternehmen zählen unter anderem Educational Testing Service,

Pearson, WriteToLearn und Gradr.<sup>1</sup> Letztlich wird sich ein wechselseitiges Spannungsverhältnis zwischen automatisierter Erstellung und automatisierter Bewertung von Arbeiten entwickeln – vergleichbar mit dem dynamischen Wechselspiel zwischen Schadsoftware und Antivirensoftware.

### 1.1.3 Mit Avataren als Ergänzung von Dozenten zum Omnichannel-Learning

Wie kann die Sinnhaftigkeit und die Motivation des Lernens gefördert werden, wenn Künstliche Intelligenz in immer mehr Bereichen bereits bessere Leistungen erbringt? Gleichzeitig entwickelt sich ChatGPT stetig weiter. Während im Jahr 2023 der Prompt, „ob Studierende ChatGPT nutzen sollten, um sich Hausarbeiten oder sogar Abschlussarbeiten schreiben zu lassen“ zu bei ChatGPT einer Verneinung führte (Dilger 2023), hat sich diese Einschätzung mit dem gleichen Prompt von ChatGPT 2025 deutlich verändert.<sup>2</sup>

Mittlerweile können aus Texten oder Vorlesungsunterlagen automatisierte Avatare als digitale Lehrassistenten erstellt werden, die im Bewegtbild mit Menschen interagieren. Solche Avatare sind bereits in der Lage, interaktiv Gespräche mit Menschen zu führen und auf ihre Fragen einzugehen. So gibt es z.B. bereits abgeschlossene Forschungsprojekte am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI), wie etwa "VoluProf"<sup>3</sup>. In diesem Projekt wurde ein interaktiver 3D-Avatar entwickelt, der mithilfe eines Sprachmodells Fragen beantworten kann. Die Interaktivität und Qualität der Antworten lässt sich weiter verbessern, wenn Professoren zusätzliche Daten, etwa aus Rankings oder Lehrmaterialien, zur Verfügung stellen.

Solche Avatare können mehr Geduld und Empathie zeigen, als es viele menschliche Lehrkräfte leisten können oder wollen. Dabei können **24/7-Avatare**, den Lernenden individuell und im eigenen Tempo begleiten und somit die Lernleistung deutlich verbessern. Derartige Avatare können im Lernprozess direkt Feedback geben und Probeausarbeitungen beliebig oft zügig korrigieren. **Automatisierte Korrekturen** bilden zusammen mit **Gamification**-Elementen und Belohnungssystemen wichtige Motivationsfaktoren beim Lernen. Da die Aufmerksamkeitsspanne abnimmt, müssen weitere Komponenten in den Lernprozess eingebunden werden. Intelligente KI-Agenten unterstützen dabei durch schnelles, permanent verfügbares personalisiertes Feedback, welches vielen Lehrkräften aufgrund begrenzter Ressourcen kaum in diesem Umfang möglich wäre.

Untersuchungen zeigen, dass die Akzeptanz von automatisiertem Essay-Scoring hoch ist, wenn es als Trainingshilfe mit schnellem Feedback eingesetzt wird. Beim Einsatz in entscheidenden Prüfungen sinkt die Akzeptanz jedoch rapide, und die Unsicherheit steigt – auch aufgrund fehlender Erfahrung und geringem Vertrauen in solche Systeme (Haried und Schneider 2025).

Damit eröffnet sich ein neuer Lernkanal. Ein **Lernkanal** ist der Weg wie sich jemand Wissen aneignen kann. Hierbei beschreibt der **Lernkanal** die bevorzugte Sinnesmodalität, über die eine Person Informationen aufnimmt und verarbeitet. In der Lernpsychologie werden typischerweise visuelle,

---

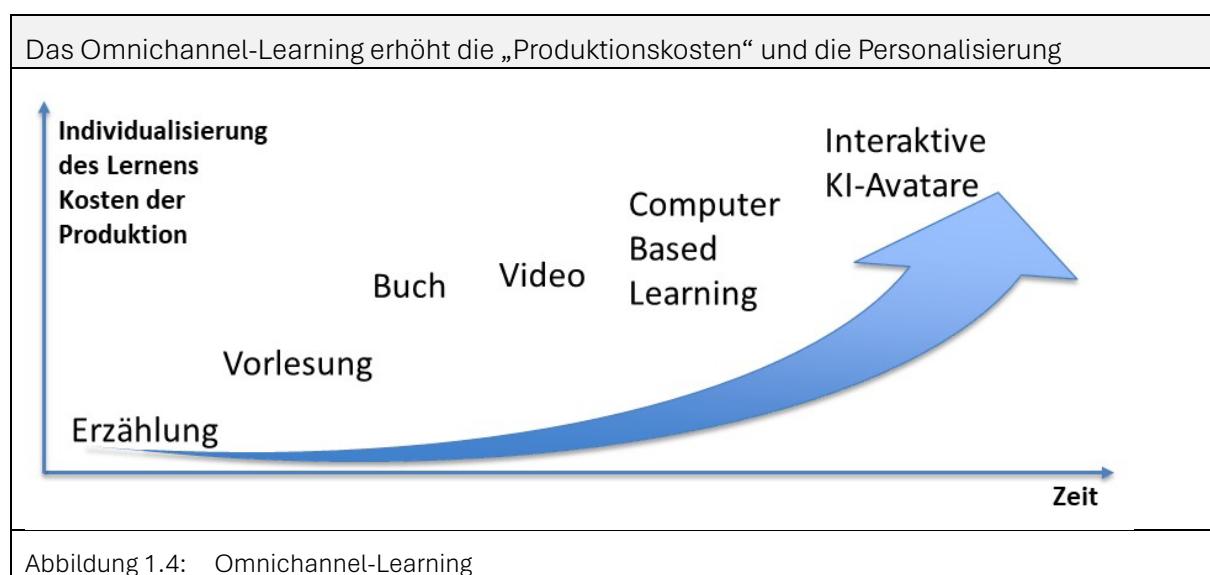
1 Siehe z.B. <https://gradr.se/product/> (07/2025)

2 Antwort von ChatGPT 03/2023: „Nein, es ist nicht ethisch oder akzeptabel, ChatGPT oder andere KI-Modelle zu nutzen, um Hausarbeiten oder Abschlussarbeiten schreiben zu lassen. Solche Praktiken werden als akademischer Betrug betrachtet und können schwerwiegende Konsequenzen haben, wie z.B. das Ablegen von Prüfungen oder sogar die Entlassung aus der Hochschule.“

Antwort von ChatGPT 08/2025: „Ja, Studierende könnten theoretisch ChatGPT oder ähnliche KI-Tools dazu nutzen, **Hausarbeiten** oder sogar **Abschlussarbeiten** zu schreiben. Tatsächlich gibt es immer mehr Berichte darüber, dass KI-generierte Texte in akademischen Umfeldern zum Einsatz kommen.“

3 Siehe [https://www.dfg.de/en/web/research/projects-and-publications/project/voluprof?utm\\_source=chatgpt.com](https://www.dfg.de/en/web/research/projects-and-publications/project/voluprof?utm_source=chatgpt.com) (08/2025)

auditive und kinästhetische Lerntypen unterschieden, wobei diese Kategorien auf bevorzugte Sinneskanäle bei der Informationsaufnahme und -verarbeitung verweisen. Die Theorie der Lernkanäle basiert auf der Annahme, dass individuelles Lernen effektiver ist, wenn es dem bevorzugten Wahrnehmungsstil entspricht (vgl. z. B. VAK-Modell in (Fleming und Mills 1992)). **Omnichannel-Lernen** ermöglicht es, Lerninhalte über verschiedene Kanäle hinweg nahtlos zu nutzen, sodass Lernende jederzeit und überall darauf zugreifen können. Dadurch wird ein flexibles, vernetztes und personalisiertes Lernerlebnis geschaffen, das sich an individuelle Bedürfnisse anpasst (Tabares und Vallejo 2020). Omnichannel bedeutet auch, dass die Lernenden nahtlos von einem Lernkanal zu einem anderen Wechseln können und wieder zurück ohne den Lernpfad verlassen zu müssen. In einer weiteren Stufe der personalisierten Unterstützung passen KI-Tools die Inhalte flexibel an Kontext und Stimmung und zur Verfügung stehender Zeit (oder benötigter Zeit) für das Lernen an. Omnichannel-Lernen ermöglicht es somit Lernenden, mit einem bevorzugten Lernkanal zu beginnen und je nach individueller Vorliebe, aktuellem Befinden oder konkreter Lernsituation flexibel zwischen verschiedenen Kanälen zu wechseln. Diese Form der Personalisierung sorgt dafür, dass das Gehirn kontinuierlich mit optimal angepasstem Input versorgt wird, was die kognitive Verarbeitung und den Lernerfolg unterstützen kann. Ein solcher **Lernprozess** ist ein zeitlich verlaufender, individueller oder sozial eingebetteter Vorgang, in dem Wissen, Fähigkeiten, Einstellungen oder Verhaltensweisen durch Erfahrung, Übung, Reflexion oder Interaktion erworben, verändert oder gefestigt wird. Dieser Lernprozess verändert sich fast nicht, da sich unser Gehirn nicht in der Geschwindigkeit der generativen KI weiterentwickelt. Nun treten zunehmend individualisierte Lernprozesse in den Vordergrund. Während früher das gesprochene Wort, die Vorlesung und später das Buch im Zentrum standen – ergänzt durch interaktive E-Learning-Systeme –, erweitert sich das Spektrum nun um einen weiteren Lernkanal: den persönlichen Lernbot (Zinn 2023). Dieser agiert nicht nur interaktiv, sondern lernt den Nutzer mit jedem Prompt besser kennen und unterstützt ihn faktenbasiert wie auch emotional im Lernprozess.



Im Rahmen von Omnichannel-Learning wird sich auch die **Rolle des Lehrenden** ändern mit einem höheren Anteil an Beratung und Begleitung von Lernenden. Der Dozent agiert als **Katalysator**, der vermehrt als **Lernbegleiter, Coach und Kurator**, als auch in der **Qualitätssicherung** in Form von Wissensüberprüfungen agiert. Der Anteil der Wissensvermittlung wird Rückläufig sein, da sich die Lernenden zunehmend auf die verschiedenen Lernkanäle aufteilen. Die Erstellung von

**Omnichannel-Lerninhalten** wird zunehmend so aufwändig, dass sie von Einzelpersonen kaum noch leistbar ist. Der Anspruch an professionelle und multimedial gestaltete Inhalte, welche mit generativer KI ergänzt wird steigt deutlich. Daher werden künftig spezialisierte akademische Content-Produzenten Materialien wie Avatare, Folien, Bücher, Videos, Podcasts oder Simulationen entwickeln, die einmal erstellt dann vielfach genutzt oder verkauft werden und ggf. von Dozenten oder Institutionen leicht angepasst werden können. Nur durch solch eine Skalierung oder durch den Einsatz automatisierter Systeme wie Avatare lässt sich der steigende Aufwand effizient bewältigen.

In diesem Kontext wird diskutiert, ob der demografische Wandel zu sinkenden Studierendenzahlen führt und die zunehmende Digitalisierung durch Omnichannel-Angebote zu Mehraufwand? Dabei rücken **lernökonomische Überlegungen** stärker in den Fokus als bisher schon. Es stellt sich die Frage, wie viele Ressourcen – in Form von Lehrenden, Technologien und Materialien pro Studierenden eingesetzt werden, welche inhaltlichen Schwerpunkte gesetzt werden sollen (z. B. Allgemeinwissen, Fachwissen oder forschendes Lernen) und welche Prüfungsformate und Gruppengrößen sinnvoll sind. Dabei geht es letztlich auch um die Frage: Welcher gesellschaftliche Nutzen wird von einem Bildungssystem erwartet? Gleichzeitig besteht die Gefahr einer **Ökonomisierung des Lernens**, bei der der Zugang zu digitalen Hilfsmitteln wie ChatGPT ungleich verteilt ist. Während einige Lernende solche Technologien selbstverständlich nutzen – teils sogar in Premium-Versionen – bleibt anderen der Zugang aus finanziellen oder anderen Gründen verwehrt. Diese Ungleichheit wirft Fragen nach Fairness und Chancengleichheit im Bildungskontext auf. So thematisiert Lordick in diesem Zusammenhang die Rolle textgenerierender KI aus schreibdidaktischer Sicht und stellt zur Diskussion, wie mit, ohne oder trotz solcher Technologien geschrieben und gelernt werden kann (Lordick 2024).

#### 1.1.4 Neue Kriterien für die Bewertung von Abschlussarbeiten

Der Einsatz von ChatGPT wird häufig mit der Einführung des Taschenrechners verglichen. Dabei stellt sich die Frage, welche grundlegenden Kenntnisse erforderlich sind, um ein solches Werkzeug sinnvoll zu nutzen – also, was man verstehen und beherrschen muss, um ChatGPT korrekt anzuwenden sowie die Ergebnisse nachzuvollziehen und erklären zu können.

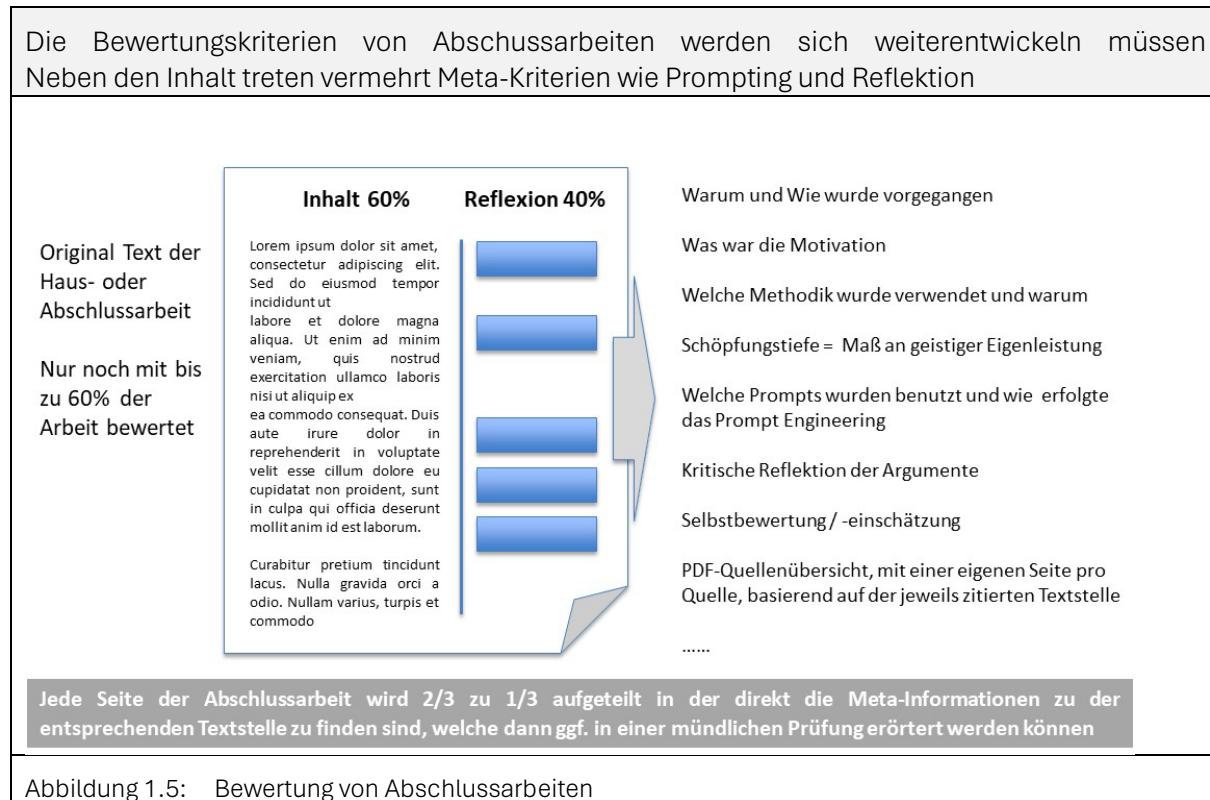
KI-generierte Texte sind aus unserem Arbeits- und Privatleben nicht mehr wegzudenken und ihr Einsatz und das Prompting muss geübt und erlernt werden. Vor diesem Hintergrund ist es keine praktikable Lösung, KI-generierte Texte als Prüfungsleistung abzuschaffen. Die zentrale Frage ist vielmehr, wie die Hochschule damit umgehen sollte und worin genau die **Eigenleistung** eines Studierenden beim Einsatz automatisierter Texterstellung besteht. Eine eindeutige Identifikation von Texten als von ChatGPT (oder ähnlichen Transformer-Modellen) erzeugt, ist nicht möglich – insbesondere nicht allein durch den Einsatz solcher Modelle selbst. Tools wie **GPTZero oder No-GPT** können allenfalls Wahrscheinlichkeiten angeben, ob ein Text KI-generiert ist. Diese Einschätzungen sind jedoch wenig zuverlässig und führen häufig zu falsch-positiven oder falsch-negativen Ergebnissen (Weber-Wulff 2023).

Bislang ist es nur eingeschränkt möglich, eine vollständige wissenschaftliche Arbeit ausschließlich durch KI erstellen zu lassen. Neben präzisem **Prompting** erfordert es weiterhin menschliche Leistungen wie Strukturierung, Gewichtung der Inhalte, Ideenfindung und vor allem die kritische Prüfung der Ergebnisse, um sogenannte „**Halluzinationen**“ – also erfundene Inhalte – auszuschließen. Auch die Überprüfung der Quellenhinweise sowie die Klärung der Originalität von Ideen müssen (bisher noch) manuell erfolgen.

Gerade hier entstehen interessante Ansätze zur Bewertung der Zusammenarbeit von Menschen und Maschine. Es handelt sich um ein typisches Beispiel eines Cyber-Physischen Systems, in dem die Qualität des menschlichen Beitrags – durch **Prompting**, kritische Reflexion, eigene Ideen oder

redaktionelle Nacharbeit – maßgeblich für die Gesamtleistung ist. Ebenso bedeutsam sind die tiefen, persönlich motivierte und verantwortete Reflexion über das Thema und die eingesetzten Methoden, die Begründung des Vorgehens und die Motivation. Auch das Verstehen und Erklären der von ChatGPT generierten Antworten, sowie die Fähigkeit diese Artefakte zu einer kohärenten Arbeit zu **orchestrieren**, sind bewertbare Leistungen. Ein sinnvoller Weg zur Sicherstellung der Eigenleistung ist eine begleitende mündliche Prüfung, in der Studierende ihr Verständnis der Inhalte, des Arbeitsprozesses und der verwendeten KI-Werkzeuge darlegen und kritisch reflektieren können.

Wie könnte eine Abschlussarbeit zukünftig aussehen und wie sollte eine Bewertung erfolgen? Das gezeigte Layout einer Abschlussarbeit in der Abbildung unten teilt jede Seite in einen 60%igen traditionellen akademischen Inhaltsteil links und einen 30%igen Reflexionsteil rechts. Der Reflexionsteil bietet Meta-Informationen zu den jeweiligen Textstellen des Inhalts, die als Grundlage für eine mündliche Prüfung dienen können.



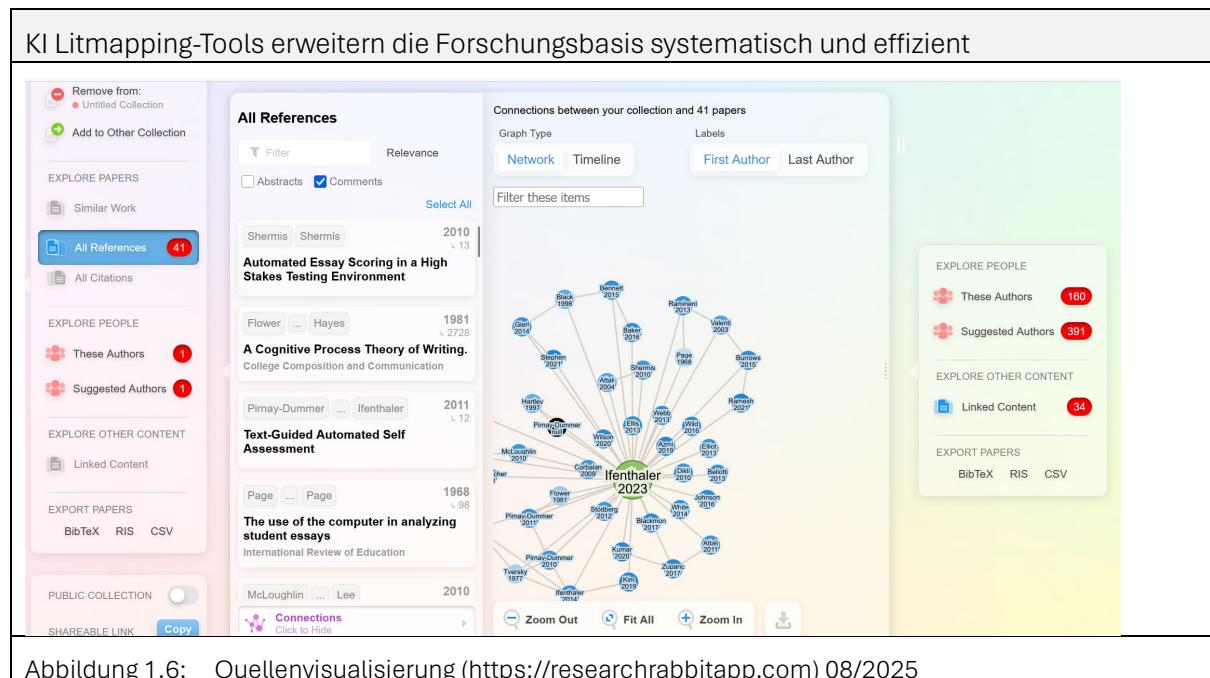
Die Meta-Kriterien einer Abschlussarbeit umfassen die Begründung des Vorgehens durch die Darlegung der Motivation und der gewählten Methodik, die Bewertung der Schöpfungstiefe als Maß an geistiger Eigenleistung inklusive der verwendeten Prompts und des Prompt Engineerings, sowie eine kritische Reflexion der Argumente, die Selbstbewertung der eigenen Leistung und eine PDF-Quellenübersicht mit einem Auszug einer Seite der Originalquelle des Zitats. Zur Sicherstellung von Transparenz und Nachvollziehbarkeit kann für jede zitierte Quelle die Seite, aus der das (direkte oder indirekte) Zitat stammt bzw. beginnt, als PDF-Auszug extrahiert und dem Anhang als ergänzendes Quellenverzeichnis beigefügt werden.

## 1.2 KI in der Forschung

### 1.2.1 Wissensproduktion im Zeitalter Künstlicher Intelligenz

KI-Tools können einerseits den traditionellen Forschungsprozess effizienter gestalten, andererseits aber auch neue Vorgehensweisen und Methoden in der Forschung ermöglichen.

Generative KI unterstützt den traditionellen Forschungsprozess durch Werkzeuge zur visuellen Darstellung relevanter Quellen eines Artikels, etwa mit Tools wie z.B. ResearchRabbit (siehe Abbildung unten). Andere unterstützende KI-Tools wie Litcit, Consensus, Ask oder Lumina Chat erleichtern die Literaturrecherche, indem sie, basierend auf Stichwörtern, einer Forschungsfrage oder einem Basisartikel, relevante und aktuelle Artikel automatisch identifizieren (Buck 2025) und kategorisieren. Wobei KI-Literaturrecherche-Tools noch Schwächen wie sprachliche Begrenzungen, eingeschränkten Zugang zu Datenquellen, Abhängigkeit vom Digitalisierungsgrad, fehlende Qualitätskontrolle, fehlerhafte Studienwiedergabe und die Illusion einer vollständigen Recherche haben.



Trotzdem transformiert generative KI die Art und Weise, wie Wissen erzeugt, verarbeitet und bewertet wird. Der Einsatz von Technologien wie ChatGPT in der Forschung führt zu fundamentalen Veränderungen wissenschaftlicher Praktiken. Prozesse werden beschleunigt, da Daten schneller analysiert und Texte automatisiert verarbeitet werden können. Es entsteht ein Wandel in den Methoden: Klassische Herangehensweisen werden durch KI-gestützte Verfahren ergänzt oder ersetzt. Gleichzeitig steigen die Anforderungen an digitale Kompetenzen, insbesondere im Umgang mit Daten, Algorithmen und KI-Systemen. Neue ethische Fragestellungen rücken in den Vordergrund, etwa hinsichtlich Transparenz, Verzerrung (Bias) und Verantwortung für KI-generierte Ergebnisse. Auch die Art und Weise, wie Wissen erzeugt und bewertet wird, verändert sich. Zugleich entstehen Abhängigkeiten von großen Technologieanbietern, während sich durch frei verfügbare Tools eine gewisse Demokratisierung der Forschung abzeichnet. Neue Herangehensweisen bieten die Chance, Forschungsprozesse nicht nur zu digitalisieren, sondern auch grundlegend zu

hinterfragen und zu optimieren – denn die bloße Digitalisierung ineffizienter Prozesse führt lediglich zu ineffizienten digitalen Abläufen.

So kann z.B. die **Blockchain-Technologie** dazu beitragen, die Nachvollziehbarkeit und Transparenz von Quellen zu gewährleisten. Obwohl der Hype um Blockchain basierende NFTs (**Non-Fungible Tokens**) nachgelassen hat, ermöglichen diese weiterhin die digitale Verfolgung von Eigentums- und Urheberrechten. Ursprungspublikationen könnten eindeutig referenziert und sämtliche Zitate sowie deren Ableitungen unwiderruflich gespeichert werden. Eine entsprechende Software könnte daraus **zentrale Zitationspfade** oder abstrahierte Wissensentwicklungen visualisieren. Generative KI-Systeme könnten auf diese Daten zugreifen, um die Herkunft, Relevanz und den Wahrheitsgehalt von Aussagen besser zu beurteilen.

## 1.2.2 Transformation in der Forschung

Es zeichnet sich ein **fundamentaler Paradigmenwechsel** ab, der sowohl Denk- als auch methodische Herangehensweisen in der Forschung grundlegend verändert. Traditionell dient das Zitieren im Sinne der Epistemologie dem Nachweis bereits vorhandenen Wissens und dem Herausstellen des Neuigkeitswerts eigener Beiträge (Popper 1959). Dies folgt dem Modell kumulativer Wissenschaftsentwicklung (Merton 1973), bei dem neues Wissen auf bestehendem aufbaut. Die klassische Vorgehensweise besteht darin, umfangreich zu recherchieren, zu lesen und den Stand der Forschung systematisch zu dokumentieren.

Mit dem Einsatz generativer KI kommt ein ergänzender, explorativer Ansatz hinzu: Forschende lassen sich zunächst Texte oder Argumentationsstrukturen erzeugen, um diese anschließend kritisch mit dem etablierten Wissensstand abzulegen. Dieser **Rückwärtsabgleich (Reverse Research)** ermöglicht eine alternative Form der Hypothesengenerierung, die sich mit dem abduktiven Schließen (Peirce 1934) verbinden lässt. Klassische Literaturrecherche und KI-gestützte Exploration sind somit komplementäre epistemologische Zugänge zur Ermittlung des Forschungsstands.

Grundlegend ist die Reflexion des Kerns der Wissensgenerierung: Zitieren macht Bekanntes kenntlich, um darauf aufbauend Neues zu schaffen und Plagiate zu vermeiden. Neue Erkenntnisse entstehen oft durch Kombination und Weiterentwicklung vorhandener Artefakte. Das Schreib- und Forschungsprozedere wird sich dabei wandeln: Statt zunächst zu recherchieren und dann selbst zu formulieren, wird die generative KI zuerst einen Textvorschlag liefern, der anschließend auf Neuheitsgehalt und Quellen geprüft wird.

Kritiker fragen, ob generative KI überhaupt Neues erschaffen kann. Hier lohnt sich ein Vergleich mit menschlicher künstlerischer Kreativität: Künstler schöpfen Neues aus ihrer Umgebung, Erfahrungen und Eindrücken, die sie durch Reisen oder Gespräche, etc. sammeln. Die KI verarbeitet Informationen aus Datenquellen wie Online-Bibliotheken oder weiteren Internetquellen und erzeugt auf Basis großer Sprachmodelle (LLMs) durch Vorhersage des nächsten Tokens neue Texte. Dadurch entstehen kreative, wenn auch datenbasierte Neuschöpfungen. Wesentlich ist dabei eine kritische Prüfung der zugrunde liegenden Trainingsdaten und -methoden. Transparenz, Zertifizierungen und nachvollziehbare Qualitätsstandards, etc. für LLMs sind unverzichtbar, um Vertrauen in die Resultate sicherzustellen.

In der bisherigen wissenschaftlichen Praxis ist/war der Mensch der zentrale Akteur – vergleichbar mit einem aktiven Spieler auf einem Fußballfeld. Mit der Einführung digitaler Werkzeuge kamen unterstützende „Bot-Mitspieler“ hinzu. Durch den Einsatz Künstlicher Intelligenz verändert sich diese Rolle jedoch weiter: Der Mensch wird zunehmend zum Trainer, der ein Team (KI-basierter) Bot-Akteure koordiniert, überwacht und strategisch einsetzt. Das Bot-Team übernimmt Aufgaben in einem Ausmaß, das zuvor kaum vorstellbar war. Das klassische rein „menschliche Spiel“ gibt es

dann (wegen Ineffektivität, Krankheit, Urlaub oder „schlafen müssen“) immer weniger - die KI arbeitet rund um die Uhr. In der Hoffnung, dass der menschliche Trainer nicht wegen Erfolglosigkeit entlassen wird.

## 1.3 Die Zukunft des Lehrens und Lernens an Hochschulen

### 1.3.1 Kommt es zu einer Hyperautomatisierung in der Hochschule?

Die zunehmende Anwendung von **Hyperautomatisierung**<sup>4</sup> an Hochschulen lässt sich als möglicher Hinweis auf einen Paradigmenwechsel im Sinne Thomas S. Kuhns (Kuhn 1962) deuten, der zu einer veränderten Weltsicht und Wissenschaftspraxis führt. Ziel ist es, sämtliche automatisierbaren Prozesse durch den Einsatz moderner Technologien wie Robotic Process Automation in der Verwaltung, Künstlicher Intelligenz zur Optimierung des Lernens, Machine Learning zur Prognose des Studienerfolgs und Process Mining zur Analyse institutioneller Abläufe grundlegend zu transformieren. Hersteller von Software im Bereich **Learning Analytics (LA)** in der Hochschulbildung beginnen mit dem Versprechen, die Lehre zu verbessern (Simbeck, 2022, S. 92). Durch die Analyse von Lerndaten soll ein tieferes und gegebenenfalls ganzheitlicheres Verständnis des Lernprozesses und der Lernenden erreicht werden.

In der Forschung kann Hyperautomation beispielsweise den Prozess der Literaturrecherche und Datenanalyse automatisieren, indem KI-Tools relevante Studien identifizieren, Daten aufbereiten und erste Auswertungen durchführen. In der Lehre können KI-gestützte Systeme individuelle Lernpfade automatisch anpassen, Feedback in Echtzeit geben und so personalisiertes Lernen fördern.

Dabei werden nicht nur organisatorische Strukturen verändert, sondern auch zentrale Annahmen über wissenschaftliche Arbeit und Lehre infrage gestellt. Besonders sichtbar wird dieser Wandel in der Nutzung generativer KI durch Studierende beim Verfassen von Abschlussarbeiten sowie automatisierter Bewertungssysteme zur Erfolgskontrolle und Prüfungsvorbereitung. Diese Entwicklungen deuten auf eine tiefgreifende Verschiebung epistemologischer und didaktischer Paradigmen hin, die als Bruch mit bisherigen wissenschaftlichen Normen im Sinne Kuhns verstanden werden kann.

Hochschulen werden ihr Wertesystem überdenken und verbindliche Leitlinien für Lehrende, Forschende und Studierende entwickeln müssen, die den Umgang mit Learning Analytics und (Generativer) Künstlicher Intelligenz regeln. Dabei sollten klare Rahmenbedingungen definiert werden, um ethische, datenschutzrechtliche und pädagogische Aspekte angemessen zu berücksichtigen. Als Orientierung könnten dabei Kriterien ähnlich den „Robotergesetzen“ von Isaac Asimov<sup>5</sup> aus dem Jahr 1942 dienen, die durch den Film I,ROBOT aus dem Jahr 2004 einem breiten

---

4 Hyperautomatisierung an einer Hochschule bedeutet den gezielten Einsatz von Technologien wie Robotic Process Automation (RPA), KI und Workflow-Management, um Verwaltungs- und Lehrprozesse – etwa Studienanmeldung, Prüfungsmanagement, Analytics oder digitale Lehre – effizienter, transparenter und weitgehend automatisiert zu gestalten.

5 Die sogenannten Robotergesetze wurden von Isaac Asimov erfunden, einem Science-Fiction-Autor. Diese fiktiven Gesetze sind:

1. Ein Roboter darf keinen Menschen verletzen oder durch Untätigkeit zulassen, dass einem Menschen Schaden zugefügt wird.

Publikum bekannt wurden und einen verantwortungsvollen und sicheren Einsatz von Technologien zum Ziel haben.

Verbindliche und nachvollziehbare Kriterien sind für eine Hochschule ebenso wichtig wie klare Prozesse und Eskalationsstufen mit menschlichen Entscheidungen	
1. <b>Transparenz, Erklärbarkeit und Interpretierbarkeit</b>	Diese Kriterien betreffen die <b>Klarheit und Nachvollziehbarkeit</b> von KI-Systemen. Es geht darum, dass Entscheidungen von KI für Menschen verständlich sind und die Funktionsweise erklärt werden kann.
2. <b>Fairness und Verzerrung (Bias)</b>	<b>Fairness</b> bedeutet die Gleichbehandlung durch KI, unabhängig von Faktoren wie Geschlecht oder Ethnie, während <b>Bias (Verzerrung)</b> unbewusste Bevorzugungen aufgrund fehlerhafter Daten beschreibt.
3. <b>Rechenschaftspflicht</b>	Entwickler und Betreiber von KI-Systemen sind für die <b>Auswirkungen</b> ihrer Technologie verantwortlich, einschließlich der ethischen und rechtlichen Konsequenzen.
4. <b>Sicherheit und Robustheit</b>	<b>Sicherheit</b> und <b>Robustheit</b> bedeuten, dass KI Fehler erkennt und in schwierigen Situationen stabil bleibt.
5. <b>Datenschutz und Schutz der Privatsphäre</b>	<b>Datenschutz</b> schützt persönliche Daten und stellt sicher, dass die Privatsphäre gewahrt bleibt und Gesetze eingehalten werden.
6. <b>Autonomie und menschliche Aufsicht</b>	Trotz der <b>Autonomie</b> von KI-Systemen sollte <b>menschliche Aufsicht</b> vorhanden sein, um unerwünschte oder schädliche Entscheidungen zu vermeiden, oder als Eskalationsstufe.
7. <b>KI-Ausrichtung auf das Wohl der Menschheit</b>	Schließlich sollten KI-Systeme so gestaltet werden, dass ihre <b>Ziele</b> mit den Interessen und dem Wohl der Hochschulmitglieder übereinstimmen, um der Gesellschaft zu dienen.

Abbildung 1.7: Kriterien zum Einsatz von KI an Hochschulen (Chan und Colloton 2024)

### 1.3.2 Entwicklung von Future Skills im Hochschulkontext

Schon früher haben sich erforderliche Fähigkeiten verändert oder wurden neu gewichtet – oft, weil Werkzeuge hinzukamen, die bestimmte Aufgaben erleichterten. Dieses Phänomen wird als "**Cognitive Offloading**" bezeichnet (Risko und Gilbert 2016). Typische Beispiele sind der Taschenrechner, der Rechenprozesse übernimmt, oder das Navigationsgerät, das uns die Wegfindung abnimmt. Auch das Schreiben kann als Mittel zur Entlastung des Gedächtnisses verstanden werden.

Zukünftig werden vor allem Fähigkeiten wichtig, die einen sicheren Umgang mit **digitalen Technologien mit sozialen und kognitiven Kompetenzen verbinden** (Klemme und Noack 2024). Neben klassischen sozialen Fähigkeiten wie Kommunikations- und Teamfähigkeit treten verstärkt technologische Fertigkeiten in den Vordergrund, etwa die routinierte Nutzung von Videokonferenzsoftware, die gemeinsame Bearbeitung zentral gespeicherter Dokumente oder das gezielte Prompting zur Interaktion mit KI-Systemen sowie die Programmierung<sup>6</sup> (**AI Literacy**).

- 
2. Ein Roboter muss den Befehlen der Menschen gehorchen, solange diese nicht dem ersten Gesetz widersprechen.
  3. Ein Roboter muss seine eigene Existenz schützen, solange dies nicht dem ersten oder zweiten Gesetz widerspricht.

6 Hierbei ändert sich die Interpretation dessen, was unter Programmierung zu verstehen ist: weg von einem rein technischen Vorgang hin zu einer strukturierten Vorgabe von Ereignissen, einer präzisen Definition von Ergebnissen sowie deren Ableitung von Konsequenzen.

Zunehmend rücken Fähigkeiten wie kritisches Denken – also das reflektierte Hinterfragen und fundierte Entscheiden – sowie digitale Kompetenzen im Umgang mit Technologien und Daten in den Vordergrund. Auch Anpassungsfähigkeit gegenüber sich wandelnden Anforderungen sowie soziale und emotionale Intelligenz gewinnen an Bedeutung. Ein zentrales Element dabei ist **Data Literacy** bzw. Datenkompetenz: Sie umfasst die Fähigkeit, Daten kritisch zu erfassen, zu verwalten, zu analysieren und gezielt anzuwenden, um informierte Entscheidungen zu treffen und überlegte Handlungen abzuleiten (Schüller 2021; Ridsdale 2015). Da Daten in nahezu allen Lebens- und Arbeitsbereichen präsent sind, gehört ein grundlegendes Verständnis datenbasierter Entscheidungsprozesse heute zur Allgemeinbildung.

In diesem Kontext entstehen neue Lern- und Prüfungskulturen: Auch entwickelt sich daraus zusätzliche Lernkanäle, die sowohl produktive als auch bewertende Automatisierung in den Hochschulalltag integriert.

## 1.4 **Fazit: Was muss man zukünftig noch Wissen, wenn ChatGPT jederzeit und überall verfügbar ist?**

Was müssen Studierende künftig wirklich wissen? Warum sollten Inhalte auswendig gelernt werden, wenn sie jederzeit gut strukturiert und ortsunabhängig über Smartphone, Wikipedia, Google oder ChatGPT abrufbar sind? Kurzfristig mag Auswendiglernen für das Bestehen von Klausuren sinnvoll sein. Doch warum überhaupt noch Klausuren? Oft wird nur das Nötigste gelernt und schnell wieder vergessen.

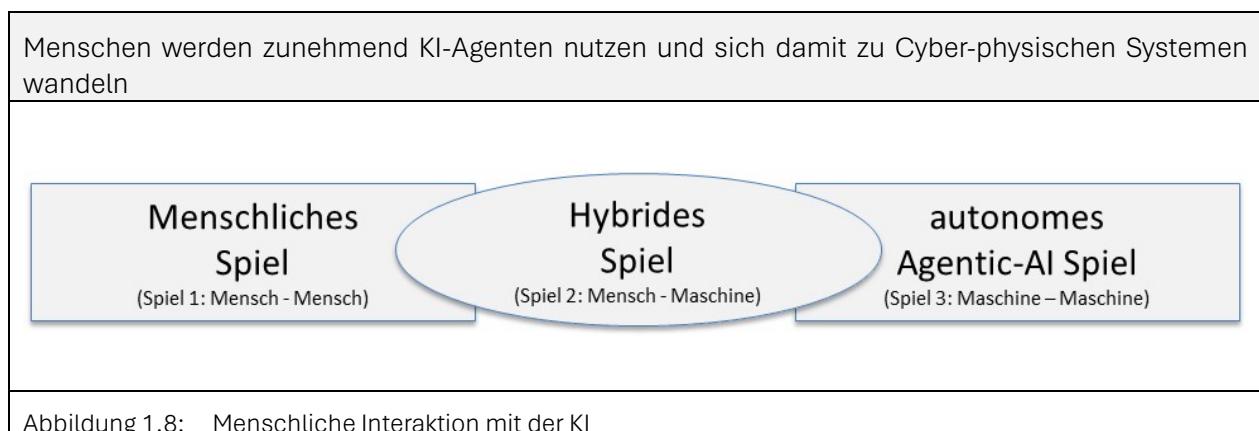
Doch nur wenn Zusammenhänge gelernt und verstanden werden, bleibt zumindest ein Überblick langfristig im Gedächtnis<sup>7</sup>, sodass Details später erneut recherchiert werden können. Wichtige Zusammenhänge sind für Gespräche und Argumentationen unerlässlich, da sie den Überblick schaffen und Fachgebiete vernetzen. Entscheidend bleiben daher Kernkompetenzen wie das präzise **Formulieren von Prompts**, der kritische Umgang mit KI sowie die Fähigkeit zum sinnvollen Austausch mit anderen.

Da Faktenwissen zunehmend maschinell verfügbar ist, gewinnen soziale Intelligenz, Werte und Normen an Bedeutung. Zudem bleiben aber auch **Fachbegriffe**, die Anwendung fachspezifischer Methoden und ein überblickartiges Verständnis des Fachgebiets unerlässlich. Die **Umsetzungskompetenz** gewinnt ebenfalls zunehmend an Bedeutung, da viele gute Ideen vorhanden sind, jedoch nur wenige in der Lage sind, diese effektiv zu realisieren

Ähnlich wie beim High Frequency Trading (HFT) bei Banken, bei dem KI-Agenten autonom (ohne menschliche Interaktion) Aktien kaufen und verkaufen wird es vermehrt Spielfelder geben, die völlig automatisiert handeln, was z.B. beim Aktienhandel schon über 70% aller Aktienkäufe weltweit durch automatisierte Agenten ausmacht (Nimalendran 2024). Es wird weiterhin menschliche Spielfelder geben z.B. Comedy vor Publikum oder eine „normale“ Unterhaltung“ ohne KI-Unterstützung. Zunehmen werden aber insbesondere hybride Spielfelder bei denen Menschen mit Unterstützung generative KI Tätigkeiten verrichten, Produkte erstellen, Analysen durchführen oder auch sich KI-Agenten unterstützt mit Avataren oder anderen Menschen unterhalten.

---

7 Siehe hierzu die Ebbinghaus'sche Vergessenskurve (1885) nach der nach 31 Tagen ca. 80 % des Gelernten wieder vergessen werden. Bei bedeutungsvollem Wissen oder regelmäßiger Anwendung (Wiederholungen) ist der Abfall allerdings deutlich geringer. Bedeutet aber auch 20% bleibt gelernt (Murre und Dros 2015).



Beim menschlichen Spiel erkennt der Gesprächspartner schnell das fachliche Know-how des Gegenübers, da keine Zeit bleibt, den KI-Agenten zu Rate zu ziehen (Spiel 1). Faktenwissen bleibt auch wichtig, um im hybriden Zusammenspiel mit generativer KI präzise Prompts zu formulieren und Ergebnisse zu bewerten (Spiel 2). Dennoch wird die Bedeutung von Faktenwissen im Vergleich zu sozialem Wissen abnehmen. Lernen wird zunehmend zum Selbstzweck – ähnlich wie viele heute ins Fitnessstudio gehen, nicht um eine bestimmte Strecke zu laufen, sondern um dem Körper Gutes zu tun. Möglicherweise wird geistige Aktivität künftig primär im Kontext sozialer Interaktion oder als kognitive Übung betrieben (Gehirnjogging<sup>8</sup>) – während viele denkbasierte Leistungen zunehmend von technischen Systemen übernommen werden (Spiel 3).

Künstliche Intelligenz ist **unerlässlich!** Echte Intelligenz aber auch!

→ Dieser Text wurde mit ChatGPT und Google Gemini orchestriert (hybrides Spiel).

## 1.5 Literaturverzeichnis

- Araujo, T., Helberger, N., Kruikemeier, S. et al. (2020) In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Soc* 35, 611–623;  
<https://doi.org/10.1007/s00146-019-00931-w>
- Buck, I. (2025) Wissenschaftliches Schreiben mit KI, DOI:  
<https://doi.org/10.36198/9783838563657>
- Chan, C; Colloton, T (2024) Generative AI in Higher Education, DOI 10.4324/9781003459026
- Dilger, A. (2023) Interview with ChatGPT about AI at Universities, <https://nbn-resolving.de/urn:nbn:de:hbz:6-50099613917>
- Fleming, N. D., & Mills, C. (1992) Not Another Inventory, Rather a Catalyst for Reflection.  
<https://doi.org/10.1002/j.2334-4822.1992.tb00213.x>
- Haried, P. Schneider, S. (2025) Artificial Intelligence in Higher Education – An Empirical Study Examining Automated Decision Making, Trust and Anxiety; in Journal of Business and Educational Leadership Vol 15, No 1: Spring 2025

<sup>8</sup> Siehe <https://digital-magazin.de/von-ainer-artgerechten-haltung-und-intelligenzgedrosselter-ki/> (11/2024)

- Ifenthaler, D. (2023). Automated Essay Scoring Systems. In: Zawacki-Richter, O., Jung, I. (eds) Handbook of Open, Distance and Digital Education. Springer, Singapore. [https://doi.org/10.1007/978-981-19-2080-6\\_59](https://doi.org/10.1007/978-981-19-2080-6_59)
- Klemme, L. & Noack, M. (2024). Kompetenzen für morgen – Diese Future Skills suchen Unternehmen schon heute. Bertelsmann Stiftung. Verfügbar unter: <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/kompetenzen-fuer-morgen>
- Kuhn, T. (1962) The Structure of Scientific Revolutions
- Lordig, N. (2024) Schreiben mit, ohne oder trotz textgenerierender Technologien? Impulse aus schreibdidaktischer Perspektive; in: Salden, P, Leschke, J. (2024) Learning Analytics und Künstliche Intelligenz in Studium und Lehre; <https://doi.org/10.1007/978-3-658-42993-5>
- Merton, R. K. (1973). The sociology of science: Theoretical and empirical investigations. Chicago: University of Chicago Press.
- Murre, J. M. J., & Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. PLOS ONE, 10(7), e0120644. <https://doi.org/10.1371/journal.pone.0120644>
- Nimalendran, M. (2024). High-frequency trading in the stock market and the costs of market liquidity. Journal of Financial Economics, 142(3), 1023–1045. <https://doi.org/10.1016/j.jfineco.2024.02.004> ScienceDirect
- Paaß, G. (2023) Foundation Models for Natural Language Processing; Springer Nature, DOI <https://doi.org/10.1007/978-3-031-23190-2>
- Peirce, C. S. (1934). Collected Papers of Charles Sanders Peirce (Vol. 5: Pragmatism and Abduction, §§ 189–212). Cambridge, MA: Harvard University Press.
- Popper, K. R. (1959). The logic of scientific discovery. London: Hutchinson.
- Ramesh, D., Sanampudi, S.K. (2022) An Automated Essay Scoring Systems: a systematic literature review. Artif Intell Rev 55, 2495–2527. <https://doi.org/10.1007/s10462-021-10068-2>
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelly, D., Matwin, S. & Wuetherick, B. (2015). Strategies and best practices for data literacy education: Knowledge synthesis report. Dalhousie University. <https://doi.org/10.13140/RG.2.1.1922.5044>
- Risko, E. F. & Gilbert, S. J. (2016). Trends in Cognitive Sciences, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Schneider, S., Haried P. (2024) Use of Artificial Intelligence in Higher Education with Particular Reference to Automated Content Generation, Trust and Anxiety, in: Journal of Business & Educational Leadership, 2024, Vol 14, Issue 1, p4, ISSN 1948-6413
- Schüller, K., Koch, H.; Rampelt, F. (2021). Data-Literacy-Charta. Version 1.2. Stifterverband; <https://www.stifterverband.org/charter-data-literacy>
- Simbeck, K. (2022). Künstliche Intelligenz und Fairness im Bildungskontext. In D. Verständig, C. Kast, J. Stricker & A. Nürnberg (Hrsg.), Algorithmen und Autonomie (S. 91–100), <https://doi.org/10.3224/84742520>
- Tabares, M, Vallejo, P. (2020) Omnichannel for Learning, DOI: <https://doi.org/10.17230/9789587207002lch2>
- Turing, A. (1950), Computing Machinery and Intelligence, Mind, LIX (236): 433–460, doi:10.1093/mind/LIX.236.433
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltynek, T., Guerrero-Dib, J., Popoola, O., Šigut, P. & Waddington, L. (2023). Testing of Detection Tools for AI-Generated Text. <https://doi.org/10.48550/arXiv.2306.15666>
- Zinn, B. (2023) Virtual Reality, Augmented Reality und Serious Games als Educational Technologies in der beruflichen Bildung; DOI 10.25162/9783515135481

## **2 Grundlagen als Ausgangspunkt empirischer Untersuchungen**

Dieses Kapitel bündelt grundlegende Begriffsbestimmungen, die den nachfolgenden Arbeiten zugrunde liegen. Um Wiederholungen zu vermeiden und ein einheitliches Verständnis zentraler Konzepte sicherzustellen, wurden die relevanten Definitionen aus den einzelnen Artikeln in diesem Abschnitt zusammengeführt. Neben der begrifflichen Klärung enthält das Kapitel auch eine Beschreibung der verwendeten Datenbasis sowie eine systematische Darstellung der methodischen Limitationen, die für die Interpretation der Ergebnisse in allen enthaltenen Arbeiten dieses Dokuments von Bedeutung sind.

### **2.1 Begriffe im Zeitalter generativer KI**

#### **2.1.1 Künstliche Intelligenz**

Der Begriff Künstliche Intelligenz (KI) wurde im Jahr 1956 während einer Konferenz am Dartmouth College in New Hampshire erstmals verwendet. John McCarthy und Marvin Minsky werden als die Gründerväter dieses Fachgebiets angesehen und legten den Grundstein für die KI als eigenständiges Teilgebiet der Informatik (Reinhart et al., 2021, S. 17).

Allgemein gesagt beschreibt KI die Fähigkeit von Computersystemen, Aufgaben zu übernehmen, für die bisher menschliches Denken und Handeln erforderlich war. Dazu gehören beispielsweise das Lernen aus Erfahrungen und das Lösen komplexer Probleme (Russell & Norvig, 2021, S.20f.).

Es wird grundsätzlich zwischen schwacher und starker KI unterschieden. Schwache KI beschreibt Systeme, die auf die Ausführung spezifischer, klar definierter Aufgaben beschränkt sind. Im Gegensatz dazu zielt starke KI darauf ab, eine maschinelle Intelligenz zu schaffen, die nicht nur Aufgaben erledigt, sondern auch die Bedeutung ihrer Handlungen versteht (Taulli, 2022, S.4).

Die meisten der heute eingesetzten KI-Systeme werden der schwachen KI zugeordnet, zu der sowohl traditionelle als auch generative KI-Systeme gehören (Heuveline & Stiefel, 2021, S.1). Traditionelle KI-Systeme basieren auf strukturierten Daten, um Vorhersagen oder Empfehlungen zu liefern. Generative KI hingegen, kann eigenständig neue Inhalte wie Texte, Bilder oder Programmcode erzeugen (Dwivedi et al., 2023; IBM iX, 2023, zitiert nach Wecke, 2024, S.6). Grundlage dafür ist das Training mit großen Datenmengen, anhand derer Muster erkannt und neue Inhalte generiert werden (Martineau, 2023).

Das Themengebiet der KI ist äußerst vielfältig und umfasst eine Vielzahl von Technologien und Aspekten. Die vorliegende Arbeit fokussiert sich ausschließlich auf generative KI, da deren Einsatz im Kontext der Klausurvorbereitung am relevantesten ist und somit einen besonderen Schwerpunkt in der Untersuchung bildet. Aus diesem Grund erfolgt keine detaillierte Auseinandersetzung mit anderen KI-Technologien.

#### **2.1.2 Definition generative KI und Einordnung im wissenschaftlichen Arbeiten**

Generative KI ist ein Teilbereich der Künstlichen Intelligenz, der eigenständig Inhalte wie Texte, Bilder oder Code erzeugt (Wecke, 2024, S.5; Neumann et al., 2024, S.229). Systeme wie ChatGPT basieren auf Foundation Models, die Inhalte nicht nur analysieren, sondern generieren (Wecke, 2024, S.6).

Grundlage ist Deep Learning mit neuronalen Netzen, insbesondere Transformer-Modelle wie GPT-4. Diese erzeugen Sprache durch statistische Vorhersagen, ohne echtes Textverständnis oder Bewusstsein (Floridi & Chiratti, 2020, S.684).

An Hochschulen wird generative KI vor allem im Kontext des wissenschaftlichen Schreibens eingesetzt. Studien belegen, dass bereits 90% der Studierenden in Deutschland auf entsprechende Tools zurückgreift. Diese werden dabei insbesondere zur sprachlichen Überarbeitung von Texten, zur Strukturierung von Argumenten sowie zur Recherche benutzt. Auch die Erstellung von Gliederungen und die Verbesserung des Schreibstils zählen zu den häufig genannten Anwendungsfeldern. (Bitkom, 2023; Forschung & Lehre, 2025; Frankfurt UAS, 2024).

Weitere zentrale Vorteile der Nutzung generativer KI bestehen in der erheblichen Zeitersparnis beim Schreiben und dem personalisierten Feedback für ihre Fragen. (Chan und Hu, 2023, S.8)

Gleichzeitig werden in der Literatur verschiedene Herausforderungen thematisiert. Dazu zählen unter anderem die Gefahr einer zunehmenden Abhängigkeit von KI, der potenzielle Verlust eigener Denk-, Schreib- und Kreativleistungen sowie Unsicherheiten darüber, in welchem Umfang und auf welche Weise KI im Schreibprozess tatsächlich eingesetzt werden darf. (Chan und Hu, 2023, S.11f.)

Ein zentrales Problem besteht darin, dass viele Hochschulen bislang keine eindeutigen Regelungen zum Einsatz generativer KI kommunizieren. Studierende wissen daher häufig nicht, was erlaubt ist und in welchen Grenzen die Nutzung akzeptiert wird (Scibbr, 2025).

Eine aktuelle Untersuchung der University of Reading zeigt zudem, dass von KI generierte Texte von Lehrenden nicht zuverlässig erkannt werden. Die Studie kommt zu dem Ergebnis, dass sich KI-gestützte Ausarbeitungen inhaltlich und formal kaum von denen anderer Studierender unterscheiden lassen und besser bewertet wurden als nicht generierte Texte. (University of Reading, 2024). Diese Erkenntnisse werfen grundlegende Fragen zur Bewertung schriftlicher Leistungen auf, insbesondere in Bezug auf die Zuweisung individueller Eigenleistung. Vor diesem Hintergrund wird deutlich, dass Hochschulen auf die veränderten Rahmenbedingungen wissenschaftlichen Arbeitens reagieren und Regelungen für den Umgang mit generativer KI entwickeln müssen.

Dies verändert nicht nur das wissenschaftliche Arbeiten, sondern wirft viel mehr die Fragen zu Verantwortung, Transparenz und Bewertung auf (Salden & Leschke, 2024). Daher ist es wichtig, generative KI nicht nur technisch zu verstehen, sondern auch im Hinblick auf Eigenleistung kritisch zu reflektieren.

### 2.1.3 Institutionelle Regelungen und rechtlicher Rahmen

Der Einsatz generativer KI im Hochschulkontext ist ein viel diskutiertes Thema, das sowohl in der Forschung als auch in der institutionellen Praxis zunehmend Beachtung findet. Insbesondere der Einfluss von KI auf Prüfungsleistungen und die Entwicklung hochschulinterner Richtlinien steht im Fokus aktueller Studien.

In einer zweiteiligen Forschung untersuchten Smolansky et al. (2023) und Kizilcec et al. (2024) die Auswirkungen generativer KI auf Prüfungen an Universitäten in Australien, den USA und Zypern. Beide Studien zeigen, dass insbesondere Hausarbeiten und Programmierprüfungen als besonders anfällig für externe Unterstützung durch KI wahrgenommen werden. Lehrende sprechen sich deutlich für angepasste Prüfungsformate aus, die den Einsatz von KI methodisch einbeziehen und auf die Förderung von kritischem Denken abzielen. Studierende hingegen zeigen sich skeptischer, da sie einen Verlust von Kreativität und Eigenständigkeit befürchten. Die internationale Perspektive von Kizilcec et al. (2024) verdeutlicht zudem kulturelle Unterschiede in der Nutzung und Akzeptanz von ChatGPT: Studierende in Zypern nutzen generative KI seltener und stehen Prüfungsanpassungen zurückhaltender gegenüber als ihre Kommilitonen in Australien oder den USA. Beide Studien unterstreichen die Notwendigkeit, Prüfungen so zu gestalten, dass sie nicht nur Leistung messen, sondern auch evaluative Kompetenzen und einen reflektierten Umgang mit KI fördern.

Parallel dazu analysierte Christ-Brendemühl (2024) die KI-Richtlinien von 67 deutschen Hochschulen mit dem Ziel, Chancen und Herausforderungen im Umgang mit generativer KI zu identifizieren. Die Analyse zeigt, dass der KI-Einsatz in vielen Fällen grundsätzlich erlaubt ist, konkrete Prüfungsregelungen aber meist dem Lehrpersonal überlassen bleiben. Besonders hervorgehoben werden der Aufbau von KI-Kompetenzen, der Einsatz zur Lehrunterstützung und die Förderung digitaler Fähigkeiten. Gleichzeitig bestehen Bedenken hinsichtlich Plagiatsrisiken, wissenschaftlicher Ungenauigkeit und Datenschutz. Ein begleitender Workshop mit Lehrenden betonte die Bedeutung einer frühzeitigen Integration von KI-Themen in das Studium sowie die kontinuierliche Weiterbildung von Lehrkräften.

Auch die HIS-HE-Studie (2025) macht deutlich, dass der rechtliche Umgang mit generativer KI derzeit vor allem durch hochschulinterne Regelungen gesteuert wird, da bundesgesetzliche Vorgaben fehlen. Die Praxis zeigt dabei ein sehr heterogenes Bild: Einige Hochschulen sprechen klare Einschränkungen oder Verbote aus, andere erarbeiten erste Leitlinien für einen reflektierten KI-Einsatz. Häufig sind Nutzung und Zulässigkeit an Bedingungen wie Transparenz, eigenständige Leistung und wissenschaftliche Standards geknüpft. Die Entscheidung darüber liegt meist beim Lehrpersonal und orientiert sich am jeweiligen Prüfungsziel. In einzelnen Fällen existieren auch generelle Verbote, insbesondere bei prüfungsrelevanten Arbeiten wie Abschlussarbeiten.

Insgesamt wird deutlich, dass Hochschulen einen differenzierten und oft noch im Aufbau befindlichen Umgang mit generativer KI verfolgen – sowohl in Bezug auf praktische Prüfungsformate als auch hinsichtlich institutioneller Richtlinien und rechtlicher Rahmenbedingungen.

#### 2.1.4 ChatGPT

ChatGPT ist ein von OpenAI entwickeltes KI-gestütztes Sprachmodell, das auf der GPT-Architektur (Generative Pre-trained Transformer) basiert und menschenähnliche Antworten auf textbasierte Eingaben generieren kann (Ray, 2023, S.122; OpenAI, 2023).

Die Entwicklung von ChatGPT begann 2018 mit GPT-1, gefolgt von GPT-2 und GPT-3, das 175 Milliarden Parameter umfasst und auf einem umfangreichen Korpus trainiert wurde, was seine Beliebtheit unter Studierenden steigerte (Ray, 2023, S.122).

GPT-4, das aktuelle Modell, ist ein multimodales Sprachmodell, das sowohl Bild- als auch Texteingaben verarbeiten kann und dabei auf mehreren akademischen und beruflichen Benchmarks Leistungen auf menschlichem Niveau erzielt (OpenAI, 2023).

## 2.2 Datenerhebung und Stichprobe

Die nachfolgenden Abschnitte beschreiben die einheitliche methodische Grundlage, auf der die Studien dieser Arbeit basieren. Die Studien wurden im Rahmen eines kursübergreifenden Forschungsprojekts zur Rolle generativer KI in wissenschaftlichen Arbeiten durchgeführt.

### 2.2.1 Forschungsdesign und Vorgehen

In der vorliegenden Arbeit wurde ein quantitatives Forschungsdesign gewählt, welches sich zur systematischen Erhebung und Auswertung von Daten eignet.

Die Datenerhebung erfolgte über eine standardisierte Online-Umfrage mit Microsoft Forms. Die Umfrage war vom 9. bis 30. Mai 2025 zugänglich und lief über einen Zeitraum von drei Wochen. Die

Verbreitung erfolgte durch Dozenten in den relevanten Vorlesungen für Studierende der Wirtschaftsinformatik, über das Lernmanagementsystem der Hochschule "CampUAS" in den entsprechenden Kursen sowie über den WhatsApp-Gruppenkanal der Wirtschaftsinformatik. Die Teilnahme war freiwillig, anonym und ohne finanzielle Anreize. Im Anschluss wurden die Daten einer statistischen Auswertung unterzogen.

Die Grundgesamtheit dieser Untersuchung umfasst ca. 500 Studierende des Bachelorstudiengangs Wirtschaftsinformatik an der Frankfurt University of Applied Sciences. An der Befragung nahmen insgesamt 66 Studierende teil (Rücklaufquote: 13,2 %).

Der eingesetzte Fragebogen wurde im Rahmen eines hochschulweiten Forschungsprojekts gemeinsam mit mehreren Studierendengruppen konzipiert. Insgesamt umfasste er sieben thematische Abschnitte mit insgesamt 34 Fragen, wovon 7 Fragen demografische Fragestellungen abbilden und allgemeine Fragen zum Einstieg in das Thema „Prüfungen in Zeiten von KI“. Darunter Fragen zum Alter, Geschlecht, Studienfortschritt und die Selbsteinschätzung der bisherigen KI-Erfahrung.

Die meisten Aussagen wurden mithilfe standardisierter Antwortformate erfasst, insbesondere durch fünfstufige Likert-Skalen (z. B. von „trifft gar nicht zu“ bis „trifft voll zu“). Die Entscheidung für Likert-Skalen erfolgte, weil sie eine strukturierte und verständliche Einschätzung subjektiver Haltungen ermöglichen. Anders als bei einer numerischen Skala mit breitem Bewertungsspielraum (z. B. 1 bis 10) profitieren Befragte hier von klar benannten Antwortoptionen, die die Urteilsbildung erleichtern. Eine fünfstufige Skala wurde gewählt, um eine gute Balance zwischen Differenzierung und Übersichtlichkeit zu gewährleisten. Eine siebenteilige Skala hätte zwar eine feinere Abstufung ermöglicht, erschien im Kontext der vorliegenden Fragestellung jedoch nicht erforderlich.

Vor der Durchführung wurde es auf Verständlichkeit und technische Funktionalität getestet.

Diese Eingrenzung erfolgte bewusst, da angenommen wurde, dass diese Zielgruppe aufgrund ihres Studienprofils sowohl mit digitalen Technologien besonders vertraut, ist als auch regelmäßig Hausarbeiten verfasst. Durch die Fokussierung auf eine homogene Stichprobe sollte zudem die Vergleichbarkeit der Antworten erhöht und potenzielle Verzerrungen durch unterschiedliche Studienbedingungen minimiert werden.

Aufgrund der nicht-probabilistischen Gelegenheitsstichprobe ist die Repräsentativität begrenzt; dennoch liefern die Daten valide Hinweise auf zentrale Zusammenhänge im Untersuchungsfeld.

## 2.2.2 Forschungslücke

Trotz der zunehmenden Relevanz generativer KI im Hochschulbereich existieren bislang nur wenige empirische Studien, die die konkrete Nutzung solcher Technologien in der Prüfungsvorbereitung und deren Auswirkungen auf Faktoren wie Prüfungsangst analysieren. Auch die Frage, welche neuen Kompetenzen im Umgang mit KI erforderlich sind, wurde bisher kaum untersucht. Die vorliegende Studie leistet daher einen Beitrag zur Schließung dieser Forschungslücke und liefert erste Erkenntnisse über studentische Erfahrungen mit KI-gestütztem Lernen.

## 2.2.3 Erhebungsinstrument und Datenerhebung

Zur Datenerhebung wurde ein selbst entwickelter Fragebogen eingesetzt, der geschlossene Items mit fünfstufigen Likert-Skalen sowie Skalen von 1 bis 7 enthält. Der Fragebogen gliedert sich in drei zentrale Bereiche:

- Soziodemografische Merkmale (Alter, Geschlecht, Semesteranzahl)
- Erfahrungen und Einstellungen zum Einsatz von KI
- Definition und Nachweisbarkeit von Eigenleistung

## 2.3 Deskriptive Statistiken

### 2.3.1 Soziodemografische Statistiken

Zur Einordnung der Ergebnisse wurden zentrale soziodemografische Merkmale wie das Geschlecht, das Alter und die Semesterzugehörigkeit erhoben.

Das folgende Säulendiagramm (Abbildung 1) zeigt die Altersverteilung der Umfrageteilnehmer. Der Großteil der Teilnehmenden war zwischen 20 und 22 Jahre alt. Weitere weniger vertretenen Altersgruppen waren 19-Jährige (vier Personen), 24-Jährige (zehn Personen) sowie einzelne Befragte im Alter von 25 bis 28 Jahren. Zwei Personen gaben an, älter als 30 Jahre zu sein.

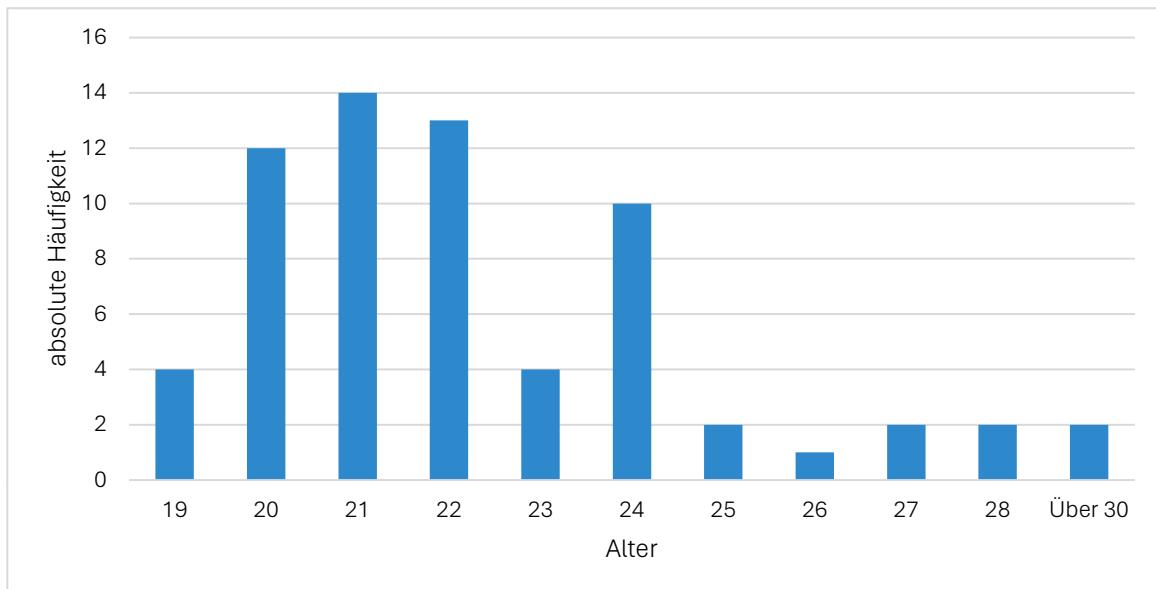


Abbildung 2.1: Altersverteilung der Umfrageteilnehmende

Abbildung 2 zeigt die Geschlechterverteilung. Die Stichprobe ( $N = 66$ ) besteht zu 68,2 % aus männlichen, zu 30,3 % aus weiblichen und zu 1,5 % aus diversen Teilnehmenden, was eine ungleiche Geschlechterverteilung mit männlichem Überhang zeigt.

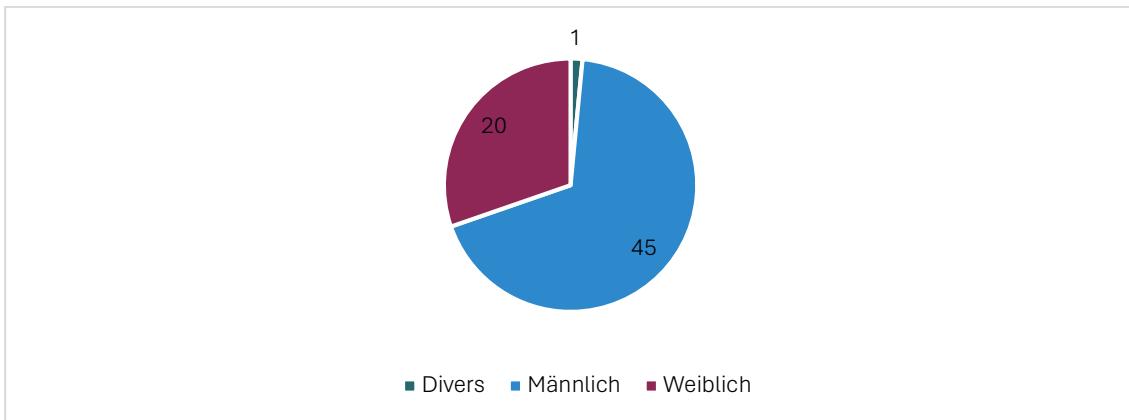


Abbildung 2.2: Geschlechterverteilung der Umfrageteilnehmende

Die Zugehörigkeit zu den Fachsemestern ist in Abbildung 3 dargestellt. Die Mehrheit studiert im 4. Semester (56,1 %), gefolgt vom 2. Semester (27,3 %). Höhere Semester (6. Semester: 10,6 %, über 8. Semester: 6,1 %) sind deutlich seltener vertreten. Die Stichprobe umfasst somit überwiegend Studierende in frühen bis mittleren Semestern.

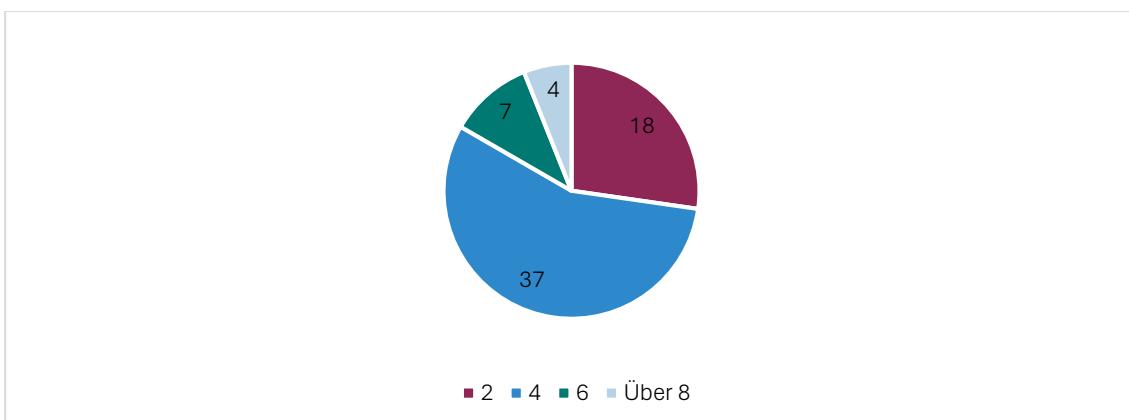


Abbildung 2.3: Semesterverteilung der Umfrageteilnehmende

### 2.3.2 Nutzung und Erfahrung mit generativer KI im Rahmen des Studiums

Neben den soziodemografischen Angaben wurden studienbezogene und KI-relevante Merkmale erhoben. Über die Hälfte der Befragten (54,5 %) verfasst nur eine Hausarbeit pro Jahr, knapp ein Drittel (31,8 %) zwei; mehr als drei Arbeiten sind die Ausnahme. Die durchschnittlich angegebene Anzahl verfasster Hausarbeiten liegt bei 1,74 pro Jahr.

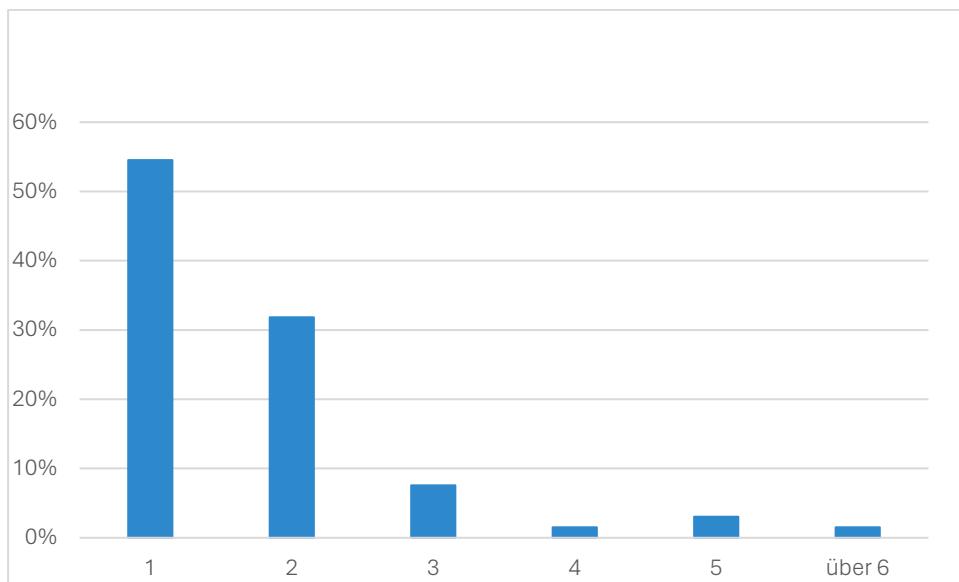


Abbildung 2.4: Anzahl an geschriebenen Hausarbeiten pro Jahr in %

Die Selbsteinschätzung zur ChatGPT-Erfahrung liegt mit einem Mittelwert von 5,45 (Skala 1–7) klar im oberen Bereich: Die meisten wählten die Stufen 5–7, lediglich vier Personen die Stufen 1–3. Demnach lässt sich ableiten, dass die meisten Teilnehmer ihre Fähigkeit mit ChatGPT umzugehen, als hoch einschätzen, was in Tabelle 1 dargestellt ist.

Bewertungsebene	Anzahl (n)	Anteil (%)
1 – Sehr gering	1	1,5 %
2	2	3,0 %
3	1	1,5 %
4	7	10,6 %
5	20	30,3 %
6	21	31,8 %
7 – Sehr hoch	14	21,2 %
Gesamt	66	100 %
Mittelwert (M)	–	5,45

Tabelle 2.1: Bewertung der eigenen ChatGPT-Erfahrung durch Studierende (N=66)

Auch die Nutzung ist hoch: 48 % verwenden ChatGPT häufig, 32 % bei jeder wissenschaftlichen Arbeit, 18 % gelegentlich und 1,5 % selten ( $M = 4,11$  auf einer 5-stufigen Skala). Die Nutzung generativer KI-Tools ist somit in allen Studienjahren fest im Studienalltag verankert, wie aus Tabelle 2 ersichtlich wird.

Nutzungshäufigkeit	Anzahl (n)	Anteil (%)
Nie	0	0 %
Selten	1	1,5 %
Gelegentlich	12	18,2 %
Häufig	32	48,5 %
Bei jeder wissenschaftlichen Arbeit	21	31,8 %
Gesamt	66	100 %
Mittelwert (M)	-	4,11

Tabelle 2.2: Häufigkeit der Nutzung von ChatGPT im universitären Kontext pro Jahr (N=66)

Abbildung 5 zeigt die Erfahrung der Studierenden im Umgang mit ChatGPT. 88% der Studierenden geben an, dass ihre Erfahrung mit ChatGPT gut bis exzellent ist. 45 Personen gaben an, männlich zu sein, 20 weiblich und 1 divers. Es fällt auf, dass 60% der männlichen Studierenden exzellente und sehr gute Erfahrungen mit ChatGPT haben, von den weiblichen Befragten haben dies jedoch nur 40%.

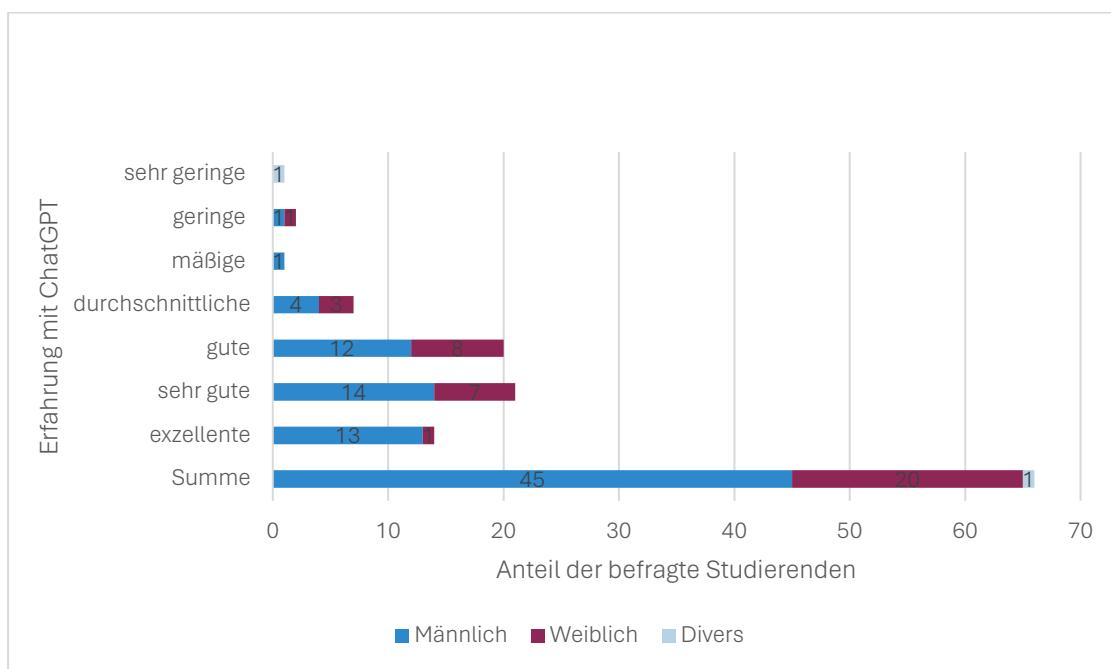


Abbildung 2.5: Einschätzung der Erfahrung der Studierenden im Umgang mit ChatGPT

Die Studierenden befanden sich zwischen dem zweiten und achten Semester ihres Bachelorstudiums in dem Fach der Wirtschaftsinformatik. Im Schnitt schrieben die Studierenden in etwa zwei akademische Arbeiten pro Jahr.

### 2.3.3 Einfluss von KI

Die Tabelle zeigt, inwieweit Studierende Aussagen zu den Auswirkungen von ChatGPT auf klassische Hausarbeiten zustimmen. Der Einsatz generativer KI wird insgesamt als unterstützend im Schreibprozess bewertet. Die höchste Zustimmung gilt der Aussage, dass KI die Qualität der Arbeit und den Arbeitsprozess verbessert. Zur Aussagekraft individueller Leistungen und zu möglichen kognitiven Beeinträchtigungen äußern sich

Aussage	Mittelwert	Standardabweichung
KI verbessert Qualität und Arbeitsprozess	4,26	0,97
KI erhöht Erwartungen	3,61	1,07
KI mindert Aussagekraft	3,30	1,16
KI beeinträchtigt kognitive Leistung	3,21	1,27

Tabelle 2.3: Einschätzung der Auswirkungen von KI auf Hausarbeiten

## 2.4 Limitationen der Studien

Dieses Kapitel benennt die Limitationen der erhobenen Daten der jeweilig betrachteten Sichtweisen.

### 2.4.1 Limitation der Arbeit: „GenKI bei Abschlussarbeiten aus Sicht der Dozenten“

Die erhobenen Daten zur Untersuchung der Sichtweise der Dozenten weist einige Limitationen auf. Die Stichprobengröße von 51 Teilnehmenden schränkt die Generalisierbarkeit der Ergebnisse ein. Zudem basiert die Erhebung auf subjektiven Einschätzungen der Lehrenden einer Fakultät, wobei lediglich zwei der vier Fachbereichen befragt wurden. Auch die verwendeten Umfrageelemente haben einen Fokus auf die Wahrnehmung von Risiken, wodurch möglicherweise positive Aspekte der KI-Nutzung nicht in ausreichendem Maße erfasst wurden.

Die Aussagekraft der vorliegenden Notendaten ist durch mehrere methodischen und kontextuellen Limitationen eingeschränkt, die ein umfassendes Verständnis der Ergebnisse erschweren. Eine zentrale Einschränkung besteht in der Nutzung aggregierter Notendaten anstelle individueller Leistungsdaten. Dadurch gehen wichtige Informationen über das konkrete Verhalten einzelner Studierender verloren. Insbesondere darüber, wie intensiv oder in welcher Form KI-Tools tatsächlich verwendet wurden. Diese Aggregation verhindert eine differenzierte Betrachtung individueller Nutzungsmuster und birgt zudem die Gefahr eines sogenannten Ecological Fallacy, bei dem fälschlich angenommen wird, dass Zusammenhänge auf Gruppenebene auch für einzelne Personen gelten.

Ein weiterer wesentlicher Schwachpunkt ist das Fehlen einer Kontrolle relevanter Störfaktoren. So konnten mögliche Veränderungen in den Prüfungsmodalitäten nicht berücksichtigt werden, etwa, ob Prüfungsformate vor oder nach der Einführung von KI angepasst wurden. Ebenso bleiben Unterschiede zwischen einzelnen Studiengängen unberücksichtigt, sowohl in der Häufigkeit als auch in der Art der KI-Nutzung. Auch qualitative Aspekte der Betreuung, etwa, wie stark Studierende durch Lehrende angeleitet oder korrigiert wurden, sowie die Komplexität der bearbeiteten Themen

könnten die Leistungen stark beeinflusst haben, ohne dass diese Variablen in die Analyse einfließen konnten.

Die Operationalisierung der sogenannten KI-Exposition stellt eine weitere methodische Begrenzung dar: Sie misst lediglich, ob KI-Tools theoretisch verfügbar waren, nicht jedoch, ob und in welchem Umfang sie tatsächlich genutzt wurden. Damit bleibt unklar, inwieweit beobachtete Effekte oder das Ausbleiben solcher, tatsächlich auf die Nutzung von KI zurückzuführen sind.

Hinzu kommen statistische Probleme, insbesondere ungleiche Gruppengrößen zwischen der Prä-KI-Gruppe (70,2 %) und der Post-KI-Gruppe (29,8 %). Diese Asymmetrie reduziert die statistische Power und erschwert den Nachweis kleiner, aber möglicherweise bedeutsamer Effekte. Auch die begrenzte Stichprobengröße und der relativ kurze Beobachtungszeitraum von lediglich vier Semestern nach der KI-Einführung schränken die Aussagekraft der Ergebnisse ein.

#### **2.4.2 Limitationen der Arbeit „GenKI bei Abschlussarbeiten aus Sicht der Studierenden“**

Die Ergebnisse der vorliegenden Untersuchung sind vor dem Hintergrund mehrerer methodischer Einschränkungen zu betrachten.

Erstens ist die Stichprobengröße mit 66 gültigen Fällen begrenzt, was die Aussagekraft statistischer Verfahren einschränkt. Die Teilnehmenden stammen ausschließlich aus dem Studiengang Wirtschaftsinformatik an der Frankfurt UAS. Dadurch ist die Übertragbarkeit auf andere Zielgruppen nur eingeschränkt möglich. Zudem kann aufgrund der freiwilligen Teilnahme eine Selbstselektion zugunsten technikaffiner oder KI-erfahrener Personen nicht ausgeschlossen werden.

Zweitens basiert die Erhebung ausschließlich auf standardisierten, geschlossenen Fragen. Dies erlaubt eine strukturierte Auswertung, begrenzt jedoch die Einsicht in individuelle Begründungen und Denkweisen. Die Antworten bleiben an den vorgegebenen Rahmen der Forschenden gebunden. Das kann zu einer gewissen theoretischen Rahmung („Framing“) der Antworten führen. Zudem basieren alle Angaben auf Selbsteinschätzungen der Teilnehmenden. Obwohl diese Methode effizient ist, birgt sie das Risiko verzerrter Angaben. Ergänzende qualitative Verfahren wie Interviews oder offene Fragen hätten zusätzliche Perspektiven eröffnet. Zudem können Selbsteinschätzungen durch soziale Erwünschtheit oder Selbstüberschätzung verzerrt sein.

Drittens ist die Aussagekraft durch die verwendeten fünfstufigen Likert-Skalen begrenzt. Diese bieten zwar eine übersichtliche Bewertung, erlauben jedoch nur eingeschränkte Differenzierung. Feinere Abstufungen oder ambivalente Haltungen lassen sich damit nur bedingt erfassen. Höher aufgelöste Skalen hätten zwar mehr Differenzierung ermöglicht, aber auch die Trennschärfe zwischen Antwortoptionen reduziert. Zudem bleibt unklar, wie intensiv oder reflektiert Teilnehmende ihre Einschätzungen getroffen haben, insbesondere bei normativen Fragen zu Prüfungsformaten oder Fairness.

Viertens ist die Umfrage hauptsächlich in Lehrveranstaltungen des 2. und 4. Semesters beworben worden. Das spiegelt sich in der Stichprobenverteilung wider. Zudem bestand die Stichprobe überwiegend aus männlich Teilnehmenden. Da sich lediglich eine Person als divers identifizierte, war eine differenzierte Auswertung dieser Gruppe nicht möglich.

Schließlich ist die zeitliche Einordnung der Erhebung im Juni 2025 zu beachten. Angesichts der dynamischen Entwicklungen im Bereich generativer KI ist davon auszugehen, dass sich sowohl die Nutzungsmuster als auch hochschulische Rahmenbedingungen innerhalb kurzer Zeit weiterentwickeln können. Die Ergebnisse stellen daher eine Momentaufnahme dar und bedürfen regelmäßiger Aktualisierung.

Trotz dieser Limitationen liefern die gewonnenen Erkenntnisse erste Tendenzen im Hinblick auf das Forschungsziel und auch eine verlässliche Basis für zukünftige Studien.

## 2.5 Literaturverzeichnis

- Bortz, J. & Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler. In Springer-Lehrbuch. <https://doi.org/10.1007/978-3-642-12770-0>
- Chan, C. K. Y. & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal Of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00411-8>
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Frankfurt University of Applied Sciences (2024): Lehren, Lernen und Schreiben mit generativen KI-Anwendungen: Empfehlungen und Impulse. Impulspapier (Stand: 11. März 2024). [online] Frankfurt University of Applied Sciences. Verfügbar unter: [https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren\\_Lernen\\_und\\_Schreiben\\_mit\\_generativen\\_KIAufwendungen\\_BeSt\\_BiB\\_Schreibwerkstatt\\_V4.2.pdf](https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren_Lernen_und_Schreiben_mit_generativen_KIAufwendungen_BeSt_BiB_Schreibwerkstatt_V4.2.pdf) [Zugriff am: 27.05.2025].
- Heuveline, V., & Stiefel, V. (2021). Künstliche Intelligenz und Algorithmen – Wahrer Fortschritt oder doch nur digitale Alchemie?
- Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S. & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Computers And Education Artificial Intelligence*, 7, 100269. <https://doi.org/10.1016/j.caeari.2024.100269>
- Martineau, K. (2023, April 20). What is generative AI? IBM Research Blog. Abgerufen am 5. Juni 2025, von <https://research.ibm.com/blog/what-is-generative-AI>
- Mhlanga, D. (2023). Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4354422>
- Neumann, K., Kuhn, J., & Drachsler, H. (2024). Generative künstliche Intelligenz in Unterricht und Unterrichtsforschung – Chancen und Herausforderungen. *Unterrichtswissenschaft*, 52(2), 227–237. <https://doi.org/10.1007/s42010-024-00212-6>
- OpenAI. (2023, 14. März). GPT-4. Abgerufen am 17. August 2023, von <https://openai.com/index/gpt-4-research/>
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Of Things And Cyber-Physical Systems*, 3, 121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- Reinhart, J., Mayer, O., & Greiner, C. (2021). Künstliche Intelligenz – Eine Einführung: Grundlagen, Anwendungsbeispiele und Umsetzungsstrategien für Unternehmen. Vogel Buchverlag.
- Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed., Global ed.). Pearson Education.
- Salden, P., & Leschke, J. (2024). Learning Analytics und Künstliche Intelligenz in Studium und Lehre: Erfahrungen und Schlussfolgerungen aus einer hochschulweiten Erprobung. Springer-Verlag.
- Smolansky, A., Cram, A., Raduescu, C., Zeivots, S., Huber, E. & Kizilcec, R. F. (2023). Educator and Student Perspectives on the Impact of Generative AI on Assessments in Higher Education. L@S '23: Proceedings Of The Tenth ACM Conference On Learning @ Scale. <https://doi.org/10.1145/3573051.3596191>
- Taulli, T. (2022). Grundlagen der künstlichen Intelligenz: Eine nichttechnische Einführung. Springer Berlin Heidelberg.

University of Reading (2024): AI-generated exam answers go undetected in real-world blind test.  
27. Juni 2024. [online] University of Reading. Verfügbar unter:  
<https://www.reading.ac.uk/news/2024/Research-News/AI-generated-exam-answers-go-undetected-in-real-world-blind-test> [Zugriff am: 14. Juni 2025].

Wecke, B. (2024). Wachstum durch den Einsatz von generativer KI: Funktionsweise und Anwendungsbereiche im Marketing (1. Aufl.). Springer Gabler.

# **Generative KI bei Abschlussarbeiten aus Sicht der Dozenten**

### **3 Der Einfluss generativer KI auf Abschlussarbeiten und Kolloquien an der Frankfurt UAS aus Sicht der Dozenten**

Marcel Crone

Philip Kreis

#### **Abstract**

Diese wissenschaftliche Arbeit untersucht den Einfluss generativer Künstlicher Intelligenz (KI) auf Abschlussarbeiten und Kolloquien an der Frankfurt University of Applied Sciences (Frankfurt UAS). Ziel ist es, die Wahrnehmung von Lehrenden hinsichtlich der Auswirkungen generativer KI systematisch zu erfassen und quantitativ auszuwerten. Im Fokus stehen Veränderungen in der Bewertungspraxis, wahrgenommene Risiken für die wissenschaftliche Integrität sowie der Zusammenhang zwischen KI-Kompetenz der Lehrenden und deren Haltung zum KI-Einsatz durch Studierende. Methodisch basiert die Arbeit auf einer Umfrage unter Lehrenden und der statistischen Analyse historischer Abschlussnoten vor und nach der flächendeckenden Verfügbarkeit generativer KI. Die Ergebnisse zeigen, dass Lehrende generative KI überwiegend als Risiko für Eigenständigkeit und Qualität studentischer Arbeiten wahrnehmen, insbesondere in Bezug auf Plagiate und sinkende Eigenleistung. Gleichzeitig wird eine Verbesserung des Schreibstils festgestellt. Statistische Analysen der Abschlussnoten belegen jedoch keinen signifikanten Einfluss der KI-Einführung auf die Bewertungsergebnisse von Thesen und Kolloquien. Die Arbeit schließt mit einer kritischen Diskussion der Ergebnisse, methodischer Limitationen und Empfehlungen für zukünftige Forschung sowie dem Umgang mit KI im Hochschulkontext ab.

**Schlagwörter:** Generative Künstliche Intelligenz (KI), Abschlussarbeiten & Kolloquien, Wissenschaftliche Integrität, Bewertungskriterien, KI-Kompetenz, Eigenständigkeit, Quellenqualität, Notenentwicklung, Transparenz

## Einordnung der Ergebnisse im Einfluss- und Veränderungsdiagramm

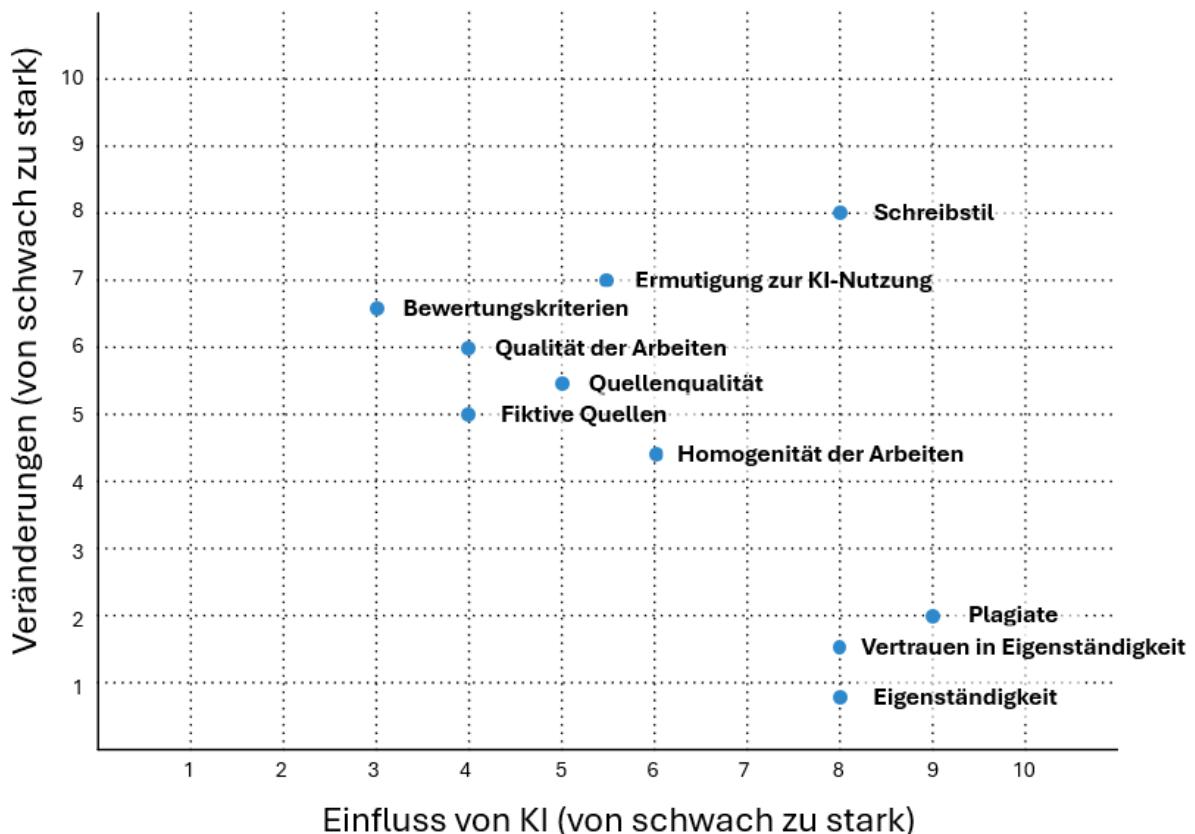


Abbildung 3.1: Zusammenhang von Veränderungen und Einfluss von KI

### 3.1 Noten unter Beobachtung: Der Einfluss von KI auf die Hochschulprüfung

Mit der zunehmenden Verfügbarkeit generativer Künstlicher Intelligenz (KI) – etwa in Form von Sprachmodellen wie ChatGPT – verändert sich auch die Art und Weise, wie wissenschaftliche Arbeiten erstellt, betreut und bewertet werden. Insbesondere an Hochschulen stellt sich die Frage, welche Auswirkungen diese Technologien auf die Eigenständigkeit, Qualität und Bewertung studentischer Abschlussarbeiten und Kolloquien haben. Lehrende sehen sich mit neuen Herausforderungen konfrontiert: der Bewertung möglicher KI-generierter Inhalte, der Einschätzung von Quellenqualität sowie der Anpassung eigener Kriterien und Standards.

Für die Hochschule Frankfurt University of Applied Sciences (Frankfurt UAS) stellt sich somit die Frage, ob und inwiefern sich der Umgang mit generativer KI bereits messbar auf die Bewertungspraxis ausgewirkt hat. Trotz vieler Diskussionen fehlt es bislang an empirischen Daten zur tatsächlichen Wahrnehmung und Reaktion der Lehrenden, sowie Auswertungen der tatsächlichen Prüfungsergebnisse.

#### 3.1.1 Zielsetzung der Arbeit

Ziel dieser Ausarbeitung ist es, die Wahrnehmung von Lehrenden an der Frankfurt UAS hinsichtlich der Auswirkungen generativer KI auf studentische Abschlussarbeiten und Kolloquien systematisch

zu erfassen und quantitativ auszuwerten. Dabei steht insbesondere im Fokus, ob Lehrende eine Veränderung der Bewertungsergebnisse feststellen und welche Risiken sie in Bezug auf die wissenschaftliche Integrität sehen. Zusätzlich wird untersucht, ob die Bewertungskriterien angepasst wurden und inwiefern die eigene KI-Kompetenz der Lehrenden mit ihrer Haltung gegenüber dem studentischen KI-Einsatz zusammenhängt. Explizit befasst sich die Hausarbeit mit der Forschungsfrage:

Hat die Verfügbarkeit generativer KI einen signifikanten Einfluss auf die Bewertungspraxis von Abschlussarbeiten und Kolloquien an der Frankfurt UAS?

Auf Grundlage dieser Forschungsfrage werden folgende Hypothesen formuliert:

- **Hypothese 1:** Wenn generative KI für studentische Abschlussarbeiten verfügbar ist, dann sehen Lehrende ein Risiko für die wissenschaftliche Qualität dieser Arbeiten und berichten von konkreten negativen Veränderungen.
- **Hypothese 2:** Lehrende haben infolge der Verfügbarkeit generativer KI ihre Bewertungskriterien für Abschlussarbeiten verändert.
- **Hypothese 3:** Lehrende, die generative KI im Alltag selbst nutzen und sich im Bereich KI als kompetent einschätzen, fördern den Einsatz von KI durch Studierende eher als weniger KI-affine Lehrende.“
- **Hypothese 4:** Generative KI, als Hilfsmittel für die Bearbeitung von Abschlussarbeiten, hat einen positiven Effekt auf die Ergebnisse
- **Hypothese 5:** Generative KI, als Hilfsmittel für die Bearbeitung von Abschlussarbeiten, mindert die Eigenleistung von Studenten und damit die Ergebnisse in mündlichen Prüfungen

### 3.1.2 Aufbau der Arbeit

Die Arbeit gliedert sich wie folgt: Nach der Einleitung erfolgt ein Überblick über den theoretischen Hintergrund zu generativer KI im Hochschulkontext. Anschließend werden die Methodik und das Vorgehen bei der Datenauswertung erläutert. Es folgt die empirische Analyse der Hypothesen. Abschließend werden die Ergebnisse diskutiert und Implikationen für die Lehre an Hochschulen aufgezeigt.

## 3.2 Methodisches Vorgehen zur Analyse von KI-Einflüssen auf Prüfungsprozesse

Zur Beantwortung der Forschungsfrage und der daraus abgeleiteten Hypothesen werden verschiedene Datenquellen herangezogen. Basierend auf der Datenlage werden unterschiedliche Analyseverfahren verwendet. Die verwendeten Methodiken der jeweiligen Analysen werden in diesem Kapitel beschrieben.

### 3.2.1 Methodik zur Analyse der Umfrage von Lehrenden der Frankfurt UAS

Im Rahmen einer Umfrage wurden Lehrende an der Frankfurt UAS befragt, wie Sie den Einfluss von generativer KI auf die Leistungen der Studenten in Abschlussarbeiten wahrnehmen. An der Umfrage haben 68 Personen teilgenommen, von denen 51 Personen die Umfrage vollständig beendet haben. Von den 51 Personen entfallen 23 (45,1 %) auf den Fachbereich 2 - Informatik und Ingenieurwissenschaften und 28 (54,9 %) auf den Fachbereich 3 - Wirtschaft und Recht.

Zur Analyse der Umfrage wurde ein quantitativer Forschungsansatz gewählt, um die Wahrnehmung von Lehrenden im Hinblick auf den Einfluss generativer KI auf studentische Abschlussarbeiten zu

untersuchen. Grundlage der Untersuchung war eine Umfrage, die sich aus geschlossenen Fragen sowie ergänzenden Freitextfeldern zusammensetzte. Die Auswertung der geschlossenen Fragen wurde deskriptiv durchgeführt. Darüber hinaus wurde der Chi-Quadrat-Test (Pearson, 1900, S. 157-175) eingesetzt, um mögliche Zusammenhänge zwischen der Nutzung generativer KI, der Anpassung von Bewertungskriterien und der Ermutigung zum KI-Einsatz durch Studierende zu untersuchen.

Ergänzend zu den quantitativen Analysen wurde eine strukturierende qualitative Inhaltsanalyse nach dem Ansatz von Mayring (2022, S. 49 - 107) vorgenommen, um die in den Freitextfeldern erhobenen offenen Angaben der Befragten systematisch zu kategorisieren und inhaltlich zu verdichten. Die Methoden wurden gewählt, um sowohl generalisierbare Tendenzaussagen als auch tiefergehende qualitative Einblicke in die subjektiven Wahrnehmungen und Bewertungshintergründe der Lehrenden zu identifizieren.

### 3.2.2 Methodik zur Analyse von Abschlussnoten

Die Umfrage bildet eine fundierte Grundlage, da sie auf dem Wissen und den Erfahrungen der Professorinnen und Professoren des Fachbereichs Wirtschaft und Recht (FB3) sowie des Fachbereichs Informatik und Ingenieurwissenschaften (FB2) der Frankfurt UAS basiert. Ergänzend dazu stellte das Prüfungsamt des FB3 anonymisierte Datensätze zu Prüfungsergebnissen, Thesen und Kolloquien im Zeitraum vom Wintersemester 2008/2009 bis zum Sommersemester 2025 zur Verfügung.

Zu Beginn erfolgte eine deskriptive Analyse dieser Daten, um einen Überblick über die Verteilung und Qualität der Datensätze zu gewinnen und darauf aufbauend geeignete statistische Auswertungsverfahren festzulegen (siehe Kapitel 4.2.1). Auf Basis dieser Voruntersuchungen wurden für die eigentliche Hypothesenprüfung der Mann-Whitney-U-Test sowie zur Überprüfung und Ergänzung der Ergebnisse eine Weighted-Least-Squares-Regression (WLS-Regression) durchgeführt.

Der Mann-Whitney-U-Test wurde gewählt, da er als nicht-parametrisches Verfahren keine Normalverteilung der Daten voraussetzt und besonders robust gegenüber unterschiedlichen Stichprobengrößen und Varianzhomogenität ist (Field, 2018, S. 443ff, Abbildung 3). Dies ist bei Notendaten essenziell, die häufig nicht-normalverteilt sind und Ausreißer enthalten können. Der Test vergleicht die Rangplätze aller Beobachtungen und ist damit sensitiv für Verteilungsunterschiede zwischen den Gruppen (McKnight & Najab, 2010, S. 1ff).

Zur Kontrolle studiengangspezifischer Effekte wurde eine WLS-Regression durchgeführt (Wooldridge, 2019, S. 279-281). Das Modell berücksichtigt die aggregierte Datenstruktur durch Gewichtung der Beobachtungen und kontrolliert für systematische Unterschiede zwischen Studiengängen durch Dummy-Variablen. Diese Spezifikation ermöglicht eine präzisere Schätzung des KI-Effekts als rein verteilungsbasierte Verfahren und die Isolierung des KI-Einflusses bei gleichzeitiger Kontrolle zeitkonstanter Störfaktoren.

## 3.3 Was Lehrende wirklich denken: KI in Abschlussarbeiten – empirische Einblicke aus Umfrage und Datenanalyse

Dieses Kapitel befasst sich mit der Vorstellung der durch die Analyse gewonnenen Erkenntnisse. Hierbei erfolgt eine separate Betrachtung der Umfrage als auch der historischen Abschlussnoten von Studierenden.

### 3.3.1 Auswertung der Umfrage von Lehrenden der Frankfurt UAS

Zur Prüfung der aufgestellten Hypothesen (1 – 3) werden unterschiedliche Antworten der Umfrage betrachtet und in Relation gesetzt. Die Ergebnisse der Umfrage und Einordnung werden in diesem Kapitel aufgeführt.

#### 3.3.1.1 Prüfung der Risiken und Auswirkungen generativer KI auf studentische Abschlussarbeiten aus Sicht der Lehrenden

Zur Prüfung der Hypothese: „Wenn generative KI für studentische Abschlussarbeiten verfügbar ist, dann sehen Lehrende ein Risiko für die wissenschaftliche Qualität dieser Arbeiten und berichten von konkreten negativen Veränderungen“ wurden die folgenden prüfbaren Hypothesen aufgestellt.

- **Nullhypothese ( $H_0$ ):** Die Verfügbarkeit generativer KI hat keinen Einfluss auf die Wahrnehmung der Lehrenden hinsichtlich der wissenschaftlichen Qualität studentischer Abschlussarbeiten, und es werden keine negativen Veränderungen berichtet.
- **Alternativhypothese ( $H_1$ ):** Die Verfügbarkeit generativer KI führt dazu, dass Lehrende ein Risiko für die wissenschaftliche Qualität studentischer Abschlussarbeiten sehen und über konkrete negative Veränderungen berichten.

Zur Überprüfung dieser Hypothesen wurden verschiedene Umfrageelemente herangezogen. Dadurch sollen verschiedene Perspektiven der Lehrenden auf Veränderungen der wissenschaftlichen Qualität von Abschlussarbeiten seit Einführung generativer KI beleuchtet werden. In Abbildung 2 wird deutlich, dass 35 der 51 befragten Teilnehmer eine Verbesserung des Schreibstils festgestellt haben. Gleichzeitig äußerten 30 Personen den Verdacht auf eine Zunahme von Plagiaten. Während 16 angaben, häufiger identische Arbeiten vorzufinden. Nur 8 Befragte nahmen eine generelle Verbesserung der Qualität von Abschlussarbeiten wahr.

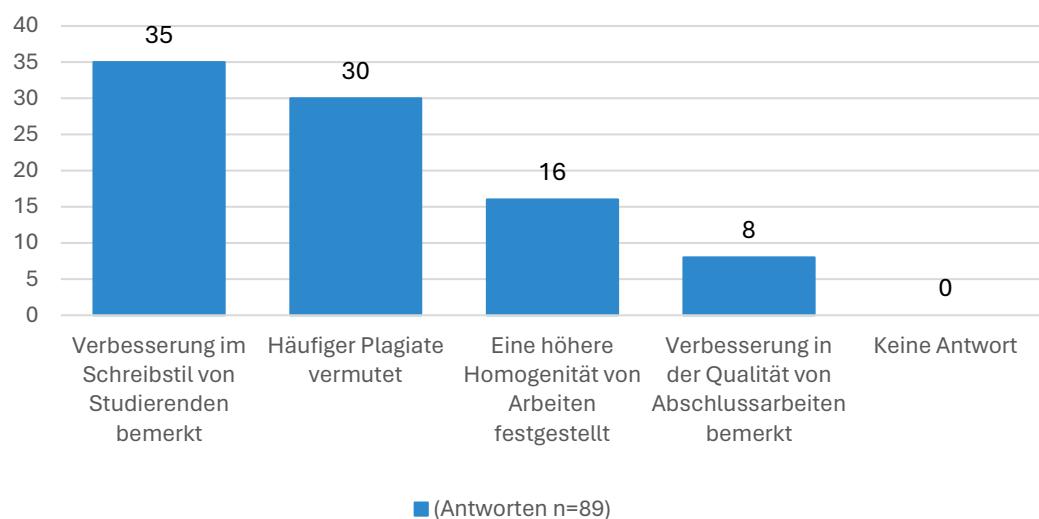


Abbildung 3.2: Wahrgenommene Veränderung seit der Verfügbarkeit von KI

Ein weiterer Bestandteil der Umfrage betrifft die Veränderung der intellektuellen Leistung, die zum Verfassen einer Abschlussarbeit erforderlich ist (Abbildung 3). Es gaben 41,18 % der Befragten an, dass diese Leistung abgenommen habe. Wohingegen 56,86 % meinten, sie sei gleichgeblieben – jedoch in anderer Form. Nur eine Person äußerte, dass die Leistung zugenommen habe.

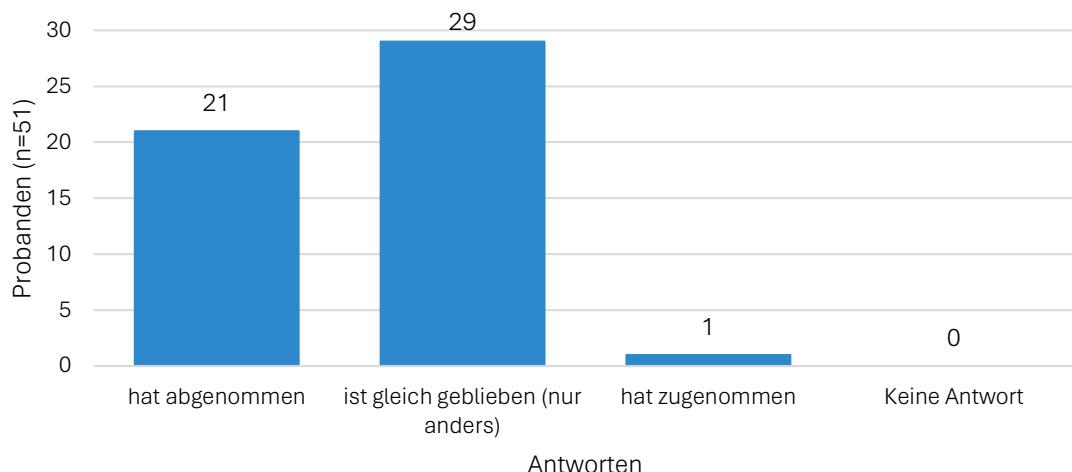


Abbildung 3.3: Veränderung der intellektuellen Leistung, um eine Abschlussarbeit zu schreiben

Auch das Vertrauen in die Eigenständigkeit der Arbeiten wurde kritisch beurteilt, wie in Abbildung 4 dargestellt. 60,78 % (n = 31) der Befragten sehen einen Rückgang studentischer Eigenleistung. Besonders ausgeprägt war dies bei jenen, die bereits zuvor eine abnehmende intellektuelle Leistung feststellten. Auch ein Teil jener, die keine Leistungsminderung wahrnahmen, gaben dennoch ein sinkendes Vertrauen in die Eigenständigkeit studentischer Leistungen an.

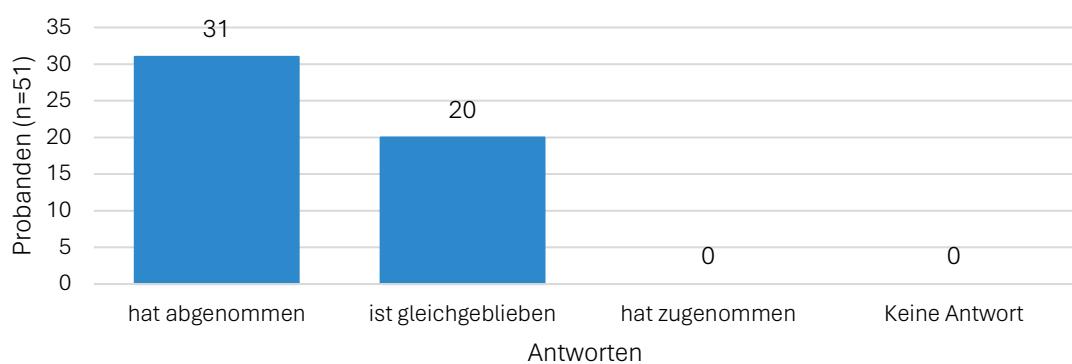


Abbildung 3.4: Vertrauen auf die Eigenständigkeit studentischer Abschlussarbeiten

Zusätzlich wurde erhoben, ob Lehrende seit der Verfügbarkeit generativer KI eine Veränderung der Quellenqualität in studentischen Abschlussarbeiten wahrnehmen. 56,86 % der Befragten (n = 29) stellten keine oder nur selten eine Veränderung fest (Abbildung 5). Demgegenüber stehen 43,14 % Teilnehmer (n = 22) die Gelegentliche bis häufige Veränderungen wahrnehmen.

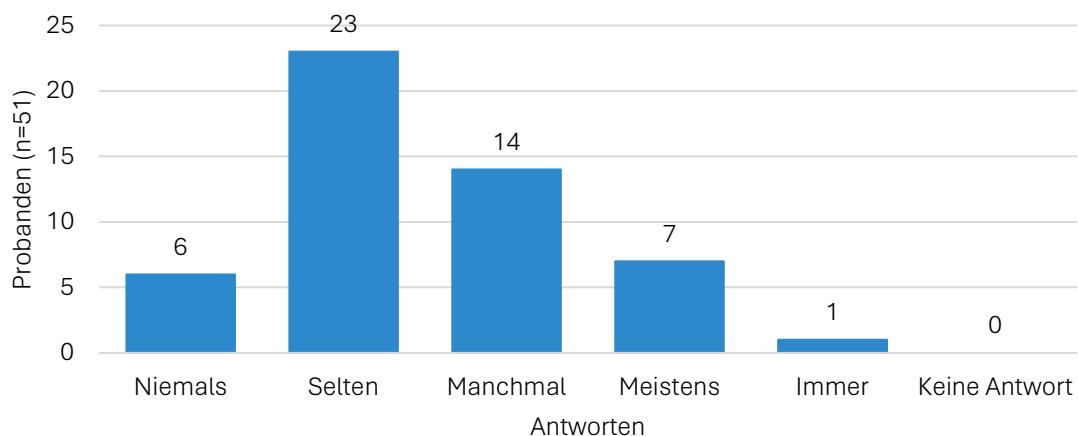


Abbildung 3.5: Veränderung der Qualität der Quellen in Abschlussarbeiten seit der Verfügbarkeit von generativer KI

In der Detailbetrachtung der Abbildung 6 wird ersichtlich, dass bei einem Teil der Lehrenden eine gestiegene Wahrnehmung von fiktiven oder falschen Quellen vorliegt. Wohingegen andere Teilnehmer, trotz vorheriger Bedenken, keine solchen Anzeichen wahrnahmen. Ein Rückschluss auf eine gestiegene Wahrnehmung von fiktiven oder falschen Quellen in Zusammenhang mit generativer KI lässt sich daher nicht ziehen.

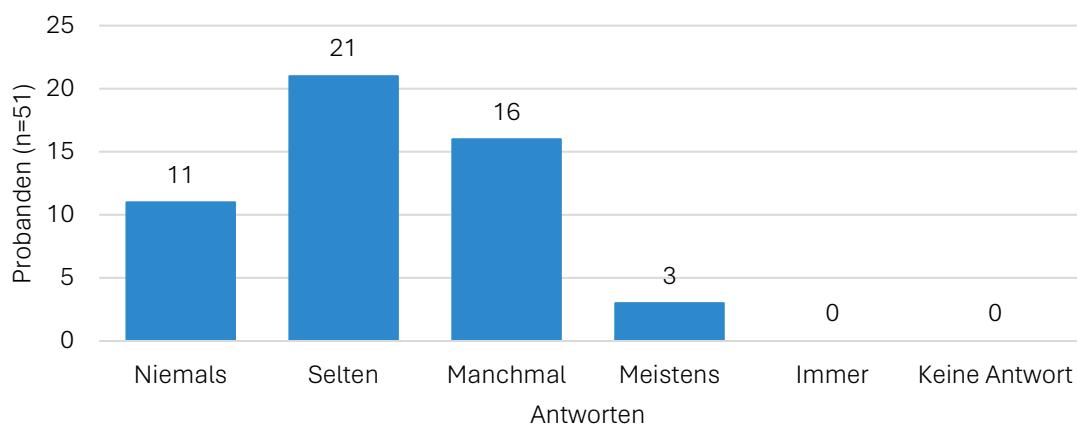


Abbildung 3.6: Wahrnehmung von fiktiven oder falschen Quellen seit der Verfügbarkeit von generativer KI

Die Ergebnisse weisen insgesamt auf ein klares, von den Teilnehmern wahrgenommenes Risiko für die wissenschaftliche Qualität studentischer Arbeiten hin. Besonders häufig wurden Plagiatsverdacht, sinkende Eigenleistung und ein Rückgang des Vertrauens in die Eigenständigkeit genannt. Die Hypothese kann daher als bestätigt gelten.

### 3.3.1.2 Prüfung veränderter Bewertungskriterien infolge der Verfügbarkeit generativer KI

Zur Prüfung der Hypothese: „Lehrende haben infolge der Verfügbarkeit generativer KI ihre Bewertungskriterien für Abschlussarbeiten verändert“ wurden die folgenden prüfbaren Hypothesen aufgestellt.

- **Nullhypothese ( $H_0$ ):** Wenn generative KI für studentische Abschlussarbeiten verfügbar ist, dann verändern Lehrende ihre Bewertungskriterien nicht.
- **Alternativhypothese ( $H_1$ ):** Wenn generative KI für studentische Abschlussarbeiten verfügbar ist, dann verändern Lehrende ihre Bewertungskriterien.

Zur Überprüfung dieser Hypothese wurden drei geschlossene Fragen aus der Umfrage ausgewertet. Zunächst wurde erhoben, ob Lehrende eine Veränderung in ihrer Bewertungspraxis vorgenommen haben. Die Ergebnisse der Umfrage sind in Abbildung 7 dargestellt. Insgesamt gaben 62,75 % (n = 32) der Befragten an, ihre Bewertungskriterien mindestens "manchmal" angepasst zu haben. Ab dieser Schwelle konnten die Teilnehmenden aus einer Liste veränderter Maßnahmen auswählen (Mehrfachauswahl war möglich).

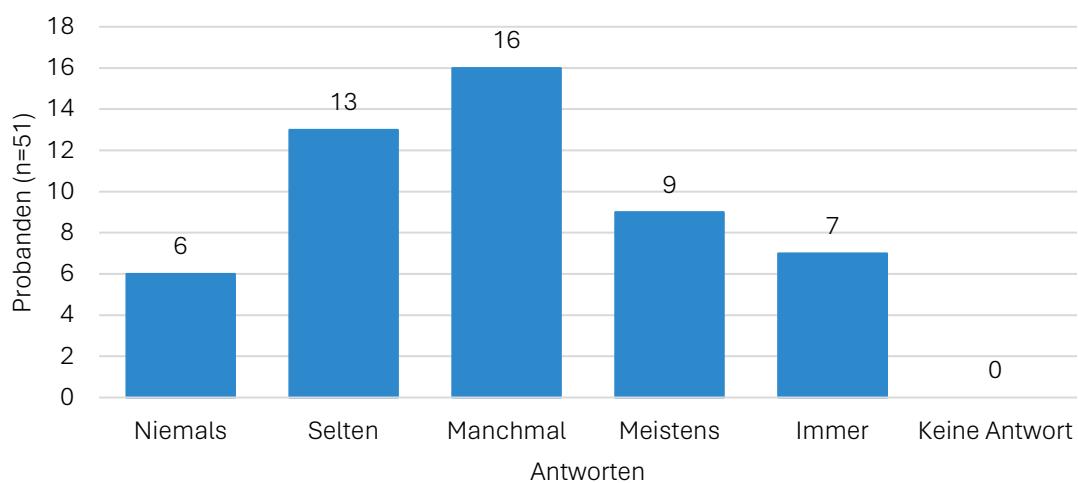


Abbildung 3.7: Veränderung der Bewertungskriterien bei Abschlussarbeiten seit der Verfügbarkeit generativer KI

Eine Gegenüberstellung der Vorgehensweise und den veränderten Bewertungskriterien ist in Tabelle 1 zu entnehmen. Besonders auffällig ist die Maßnahme „KI-Angabe verpflichtend“, die von 93,75 % der Teilnehmer gewählt wurde, die ihre Bewertungskriterien manchmal angepasst haben. Die Maßnahme „Offenlegung des Denkprozesses“ wurde dagegen häufig von jenen gewählt, die ihre Kriterien „immer“ angepasst haben (85,7 %). Jedoch weniger relevant für die Personen die „manchmal“ angegeben haben. Die „mündliche Verteidigung“ spielt hingegen in allen Gruppen eine untergeordnete Rolle.

Antwort	KI-Angabe verpflichtend	Mündliche Verteidigung	Offenlegung des Denkprozesses	Kritischere Prüfung der Abschlussarbeit
Manchmal (N=16)	15 (93,75 %)	4 (25 %)	7 (43,75 %)	9 (56,25 %)
Meistens (N=9)	6 (66,7 %)	1 (11,1 %)	5 (55,6 %)	8 (88,9 %)
Immer (N=7)	5 (71,4 %)	3 (42,9 %)	6 (85,7%)	5 (71,4 %)

Tabelle 3.1: Kreuztabelle Bewertungsmaßnahmen

Zur statistischen Überprüfung wurde ein Chi-Quadrat-Test (Pearson, 1900, S. 157-175) auf Unabhängigkeit durchgeführt. Dieser ergab einen Wert von  $\chi^2 (12) = 3,98$  mit einem p-Wert von 0,985. Da der p-Wert deutlich über dem Signifikanzniveau von 0,05 liegt, kann die Nullhypothese nicht verworfen werden.

In einem zusätzlichen Freitextfeld konnten Lehrende weitere Anpassungen ihrer Bewertungspraxis beschreiben. Diese wurden im Rahmen einer strukturierenden qualitativen Inhaltsanalyse nach Mayring (2022, S. 49 - 107) ausgewertet. Fünf thematische Kategorien konnten dabei identifiziert werden.

- **Sprachliche Qualität und Stil:** Lehrende achten zunehmend weniger auf perfekte sprachliche Formulierungen, da diese mit KI-Nutzung assoziiert werden. Stattdessen rücken Argumentationstiefe und Substanz stärker in den Fokus. Sprachliche Unregelmäßigkeiten gelten dabei zum Teil als Hinweis auf eigenständige Autorenschaft.
- **Eigenständigkeit und Kontrolle:** Um die Eigenleistung der Studierenden besser abzusichern, setzen Lehrende verstärkt auf begleitende Maßnahmen wie Kolloquien, Betreuungsintervalle und die Dokumentation des Schreibprozesses.
- **Quellenarbeit und Transparenz:** Von den Studierenden wird eine präzisere Angabe und kritische Auseinandersetzung mit Quellen gefordert, etwa durch kommentierte Bibliografien oder reflektierte Literaturanalysen.
- **Anpassung der Aufgabenstellungen:** Aufgaben werden gezielter so formuliert, dass sie individuelle Bearbeitung und originelle Lösungen erfordern – z. B. durch empirische Datenerhebungen oder anwendungsbezogene Fallbeispiele.
- **Reflektierter KI-Einsatz:** Die Nutzung generativer KI wird nicht grundsätzlich ausgeschlossen. Erwartet wird jedoch ein kritischer und transparenter Umgang damit, etwa durch Angabe der verwendeten Tools und Reflexion ihrer Rolle im Schreibprozess.

Diese qualitativen Erkenntnisse zeigen, dass sich Lehrende in ihrer Bewertungspraxis zunehmend auf individuelle, analytische und nachvollziehbare Leistungen konzentrieren. Es kann jedoch kein statistisch signifikanter Zusammenhang zwischen der Häufigkeit der Veränderung und der Auswahl konkreter Bewertungsmaßnahmen festgestellt werden. Auch wenn sich deskriptiv Unterschiede zwischen den Gruppen zeigen, reichen diese nicht aus, um eine systematische Abhängigkeit anzunehmen.

### 3.3.1.3 Prüfung der Zusammenhänge zwischen der eigenen KI-Nutzung von Lehrenden und der Unterstützung studentischer KI-Anwendung

Zur Prüfung der Hypothese: „Lehrende, die generative KI im Alltag selbst nutzen und sich im Bereich KI als kompetent einschätzen, fördern den Einsatz von KI durch Studierende eher als weniger KI-affine Lehrende.“ wurden die folgenden prüfbaren Hypothesen aufgestellt.

- **Nullhypothese ( $H_0$ ):** Es besteht kein Zusammenhang zwischen der Nutzung generativer KI im Alltag und der subjektiven Kompetenzwahrnehmung von Lehrenden und der Förderung des KI-Einsatzes durch Studierende.
- **Alternativhypothese ( $H_1$ ):** Lehrende, die generative KI im Alltag nutzen und sich als kompetent einschätzen, fördern den Einsatz von KI durch Studierende häufiger als Lehrende mit geringer Nutzung und geringem Kompetenzempfinden.

Zur Prüfung dieser Hypothese wurden die Teilnehmer gefragt, ob Sie studierende ermutigen generative KI für Abschlussarbeiten zu verwenden. Die Ergebnisse sind in der Abbildung 8 dargestellt. Insgesamt ist erkennbar, dass ein Großteil der Lehrenden ( $n = 31$ ; 60,8 %) Studierende zum mindesten gelegentlich zur Nutzung generativer KI in Abschlussarbeiten ermutigt, wobei 35,3 % ( $n = 18$ ) dies häufig tun. Dem gegenüber steht eine zweite Gruppe ( $n = 12$ , 23,5 %), die sich klar gegen eine Empfehlung ausspricht.

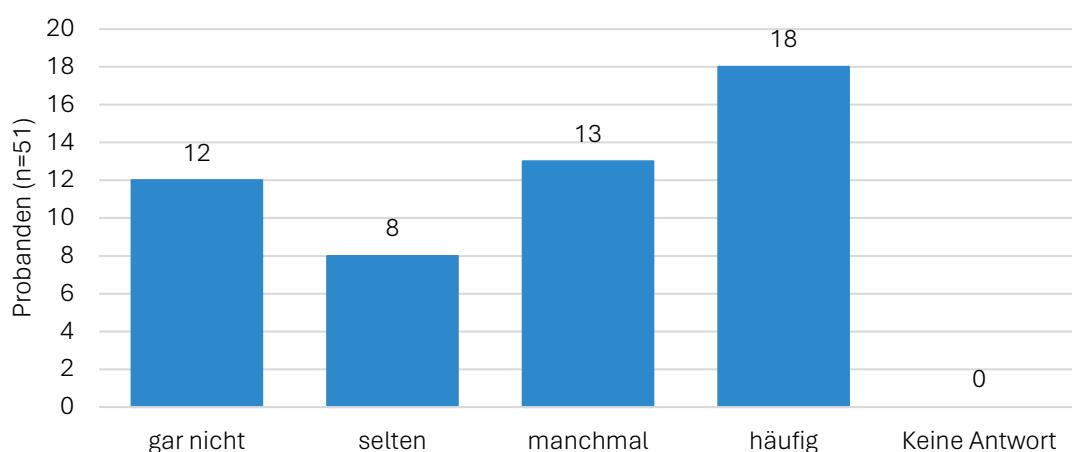


Abbildung 3.8: Ermutigung von Studierenden zum Einsatz von generativer KI in der Bearbeitung von Abschlussarbeiten

Zusätzlich wurden die Teilnehmer befragt, wie regelmäßig Sie KI im Alltag nutzen. Die Analyse der Tabelle 2 zeigt einen signifikanten Zusammenhang ( $\chi^2(9) = 26,65$ ,  $p = 0,002$ ). Lehrende, die KI häufig im Alltag nutzen, ermutigen auch häufiger Studierende. Die Nullhypothese wird in diesem Teilaспект verworfen.

		Ermutigung des KI-Einsatzes durch Studierende			
		Gar nicht	Selten	Manchmal	Häufig
Alltägliche Nutzung generativer KI	Gar nicht	2	0	0	0
	Selten	10	6	6	3
	Täglich	0	2	5	10
	Mehrfach Täglich	0	0	2	5

Tabelle 3.2 Kreuztabelle Alltägliche Nutzung von KI und Ermutigung von KI-Einsatz durch Studierende

Darüber hinaus wurden die Teilnehmer um eine Selbsteinschätzung ihres Wissens zu generativer KI gebeten. Die Tabelle 3 zeigt keinen signifikanten Zusammenhang ( $\chi^2(12) = 11,39, p = 0,496$ ). Die Empfehlungshaltung verteilt sich gleichmäßig über alle Wissensniveaus. Die subjektive Kompetenz scheint somit kein ausschlaggebender Faktor für die Ermutigung zur Nutzung von KI zu sein. Die Nullhypothese bleibt in diesem Teilespekt bestehen.

		Wie gut schätzen Sie Ihr Wissen über generative KI ein?				
		Sehr schlecht	Schlecht	Weder noch	Gut	Sehr gut
Ermutigung des KI-Einsatzes durch Studierende	Gar nicht	1	1	7	3	0
	Selten	0	1	4	2	1
	Manchmal	0	0	7	3	3
	Häufig	0	1	5	6	6

Tabelle 3.3: Kreuztabelle Ermutigung zur Nutzung von KI und Selbsteinschätzung des Wissens über generative KI

Teilnehmer, die sich gegen eine Ermutigung aussprechen oder diese selten vornehmen hatten die Möglichkeit ihre Bedenken zu äußern. Die Ergebnisse hierzu sind in Abbildung 9 abgebildet. Mehrfachangaben waren möglich. Besonders häufig wurde der Verlust des kritischen Denkens der Studierenden durch den Einsatz generativer KI ( $n = 15$ ) angegeben. Plagiate und der Rückgang der akademischen Integrität wurden von der Hälfte der für diese Frage zugelassenen Teilnehmer angegeben. Etwas weniger ausgeprägt waren die Sorgen in Bezug auf Datenschutz, Urheberrecht und Ethik ( $n = 8$ ). Weniger relevant für die Befragten waren die Veränderungen der Lehrmethodik ( $n = 3$ ) und einer unfairen Bewertung ( $n = 2$ ).

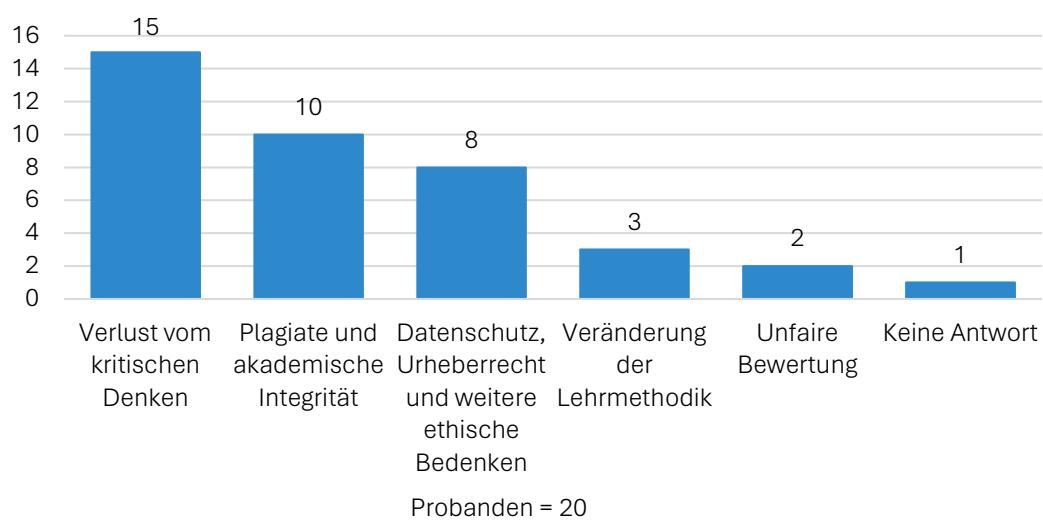


Abbildung 3.9: Bedenken bezüglich der Ermutigung zur Nutzung generativer KI in Abschlussarbeiten

Ergänzend konnten weitere Anmerkungen als Freitext angegeben werden. Diese wurden nach Mayring (2022, S. 49–107) ausgewertet. Aus der qualitativen Inhaltsanalyse lassen sich vier zentrale Kategorien ableiten:

- **Mangelnder Kompetenzerwerb:** Studierende würden sich durch KI weniger intensiv mit den Inhalten ihrer Arbeit auseinandersetzen, wodurch wichtige Fähigkeiten wie Problemlösungskompetenz und Strukturierungsvermögen nicht ausreichend ausgebildet würden.
- **Begrenzter Nutzen für wissenschaftliches Arbeiten:** Zwar könnte der Schreibstil durch KI verbessert werden, allerdings seien die generierten Texte oft inhaltsleer und zu allgemein gehalten.
- **Wissenschaftliche Standards:** Lehrende betonten die Notwendigkeit eigenständiger Auseinandersetzung mit Theorien sowie präziser, überprüfbarer Quellenangaben.
- **Unabhängigkeit der Modelle:** Es wurden ethisch-politische Bedenken bezüglich der Herkunft und ideologischen Ausrichtung von KI-Modellen geäußert, insbesondere wenn diese aus autoritären oder wirtschaftlich dominierten Kontexten stammen.

### **3.3.1.4 Auswertung der historischen Abschlussnoten**

Für die Auswertung wurden Daten des Prüfungsamts des FB3 der Frankfurt UAS verwendet, die als konsolidierte XLSX-Datei vorlagen. Die Datensätze umfassen insgesamt 10290 Notenvergaben aus dem Zeitraum Wintersemester 2008/2009 bis Sommersemester 2025. Erfasst wurden dabei für jeden Studiengang und für jedes Semester die Häufigkeit der vergebenen Noten. Die Rohdaten enthalten Angaben zu Fachbereich, Abschlussart, Studiengang, Prüfungsordnungsversion, Prüfungsnummer, Prüfungsbezeichnung, Prüfungssemester, Note und Anzahl der jeweiligen Notenvergaben. Für die Analyse wurden die Daten zunächst codiert und vereinheitlicht, beispielsweise durch die Zusammenfassung unterschiedlicher Bezeichnungen wie „Prüfung: Thesis“ und „Prüfung: Bachelorarbeit“ unter dem einheitlichen Begriff „Thesis“. Zunächst erfolgten keine weiteren Veränderungen an den Rohdaten.

### **3.3.1.5 Deskriptive Analyse der Daten**

Zu Beginn der statistischen Auswertung wurde eine deskriptive Analyse der zu Verfügung gestellten Daten durchgeführt.

Abbildung 10 zeigt ein Balkendiagramm, dass die Anzahl der erfassten Kolloquien und Thesen für jedes Semester in chronologischer Reihenfolge darstellt. Bereits bei der ersten Betrachtung fällt auf, dass in den Anfangssemestern nur sehr wenige Datenpunkte vorhanden sind. Zudem deutet die geringe Anzahl an Einträgen für das Sommersemester 2025 darauf hin, dass die Datenerfassung für diesen Zeitraum möglicherweise noch nicht abgeschlossen ist. Um Verzerrungen durch unvollständige oder fehlende Daten zu vermeiden, wird die weitere Analyse daher auf den Zeitraum vom Sommersemester 2015 (20151) bis zum Wintersemester 2024/2025 (20242) beschränkt. Abbildung 11 visualisiert die Aufstellung mit dem eingeschränkten Zeitraum. Durch die Einschränkung sinkt die Anzahl der auswertbaren Notenvergaben um 8,4 %, von 10290 auf 9425.

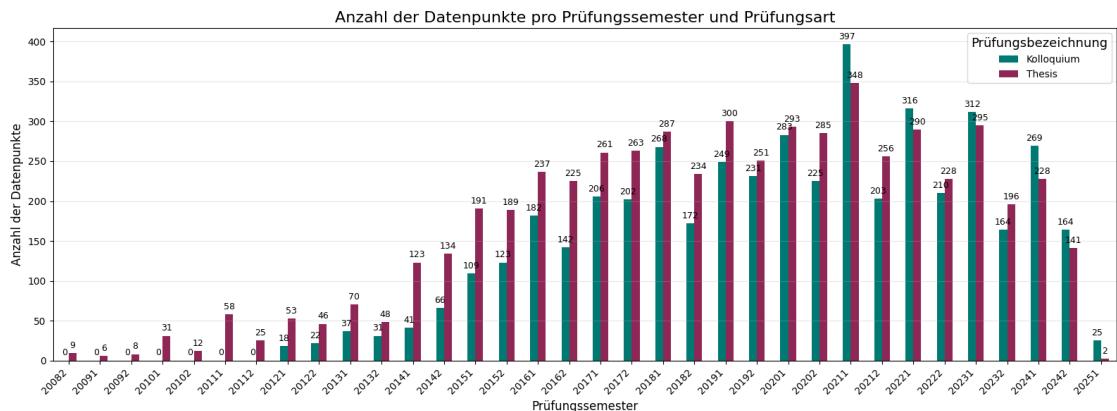


Abbildung 3.10: Anzahl der Datenpunkte pro Prüfungssemester und Prüfungsart (Vor Dateneinschränkung)

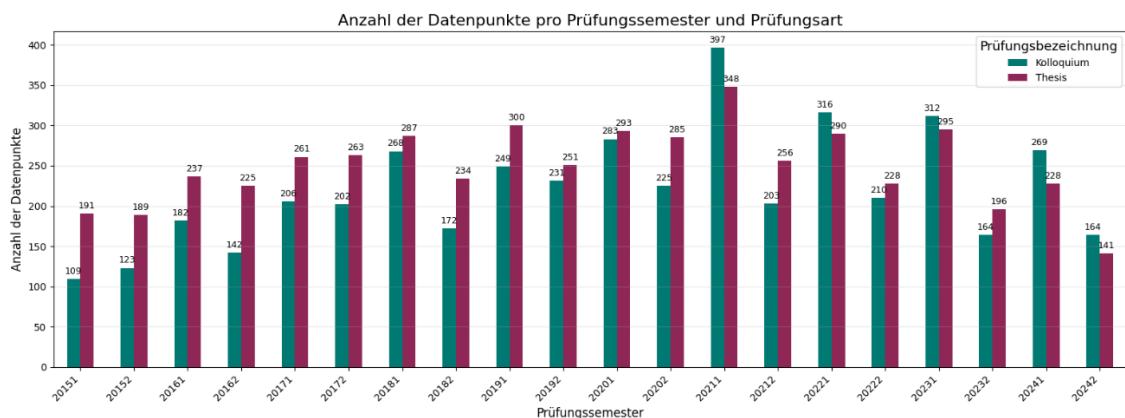


Abbildung 3.11: Anzahl der Datenpunkte pro Prüfungssemester und Prüfungsart (Vor Dateneinschränkung)

Abbildung 12 zeigt den Verlauf der durchschnittlichen Noten über die Semester hinweg sowie einen gleitenden Durchschnitt der letzten drei Jahre.

Die Daten des Prüfungsamtes wurden für die anschließende Prüfung der Hypothesen 4 und 5 (Kapitel 4.2.2 und 4.2.3) in zwei Gruppen eingeteilt: Prüfungsleistungen vor und nach der flächendeckenden Verfügbarkeit generativer KI. Der Zeitpunkt der Veröffentlichung von ChatGPT im November 2022 ist als Markierung für die flächendeckende Verfügbarkeit generativer KI im Diagramm hervorgehoben und wurde als Stichtag für diese Einteilung der Semester in „Prä-KI“ und „Post-KI“ genutzt (OpenAI, 2022). Prüfungen, die ab dem Wintersemester 2022/2023 (Prüfungssemester  $\geq 2022/1$ ) abgelegt wurden, werden daher der Gruppe „Post-KI“ zugeordnet, während alle vorherigen Prüfungen der Gruppe „Prä-KI“ zugeordnet werden. Aus der Darstellung wird deutlich, dass Kolloquien im Durchschnitt durchgehend besser bewertet werden als Thesen.

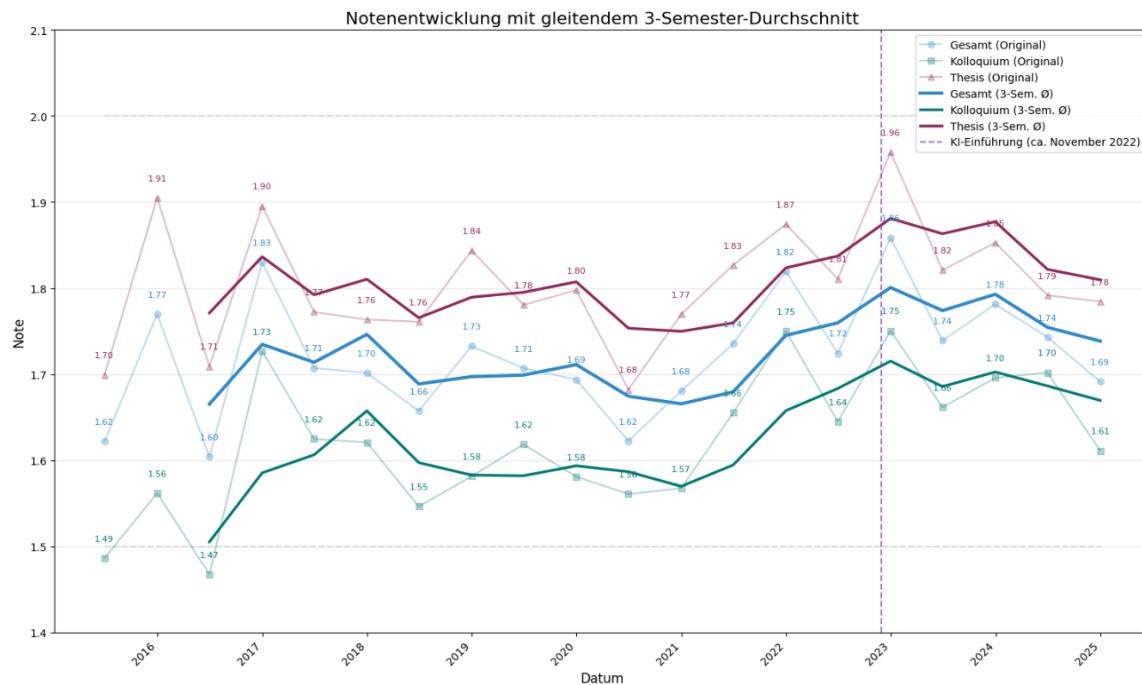


Abbildung 3.12: Notenentwicklung mit gleitendem 4-Semester Durchschnitt

Abbildung 13 visualisiert die Verteilung der Dateneinträge nach Kategorien. Das Verhältnis von Kolloquien zu Thesen ist – wie erwartet – nahezu ausgeglichen, da in der Regel zu jeder Thesis auch ein Kolloquium gehört. Deutlich ist hingegen die Diskrepanz bei der Anzahl der erfassten Einträge vor und nach der flächendeckenden Verfügbarkeit generativer KI: Während die Gruppen Kolloquium und Thesis annähernd gleich groß sind, ist der Anteil der Datensätze aus der Post-KI-Phase im Vergleich zur Prä-KI-Phase deutlich geringer. Die Einführung von KI-basierten Tools betrifft bislang nur einen Teil des Untersuchungszeitraums wodurch die Datenbasis für die Zeit nach der KI-Einführung entsprechend kleiner ist.

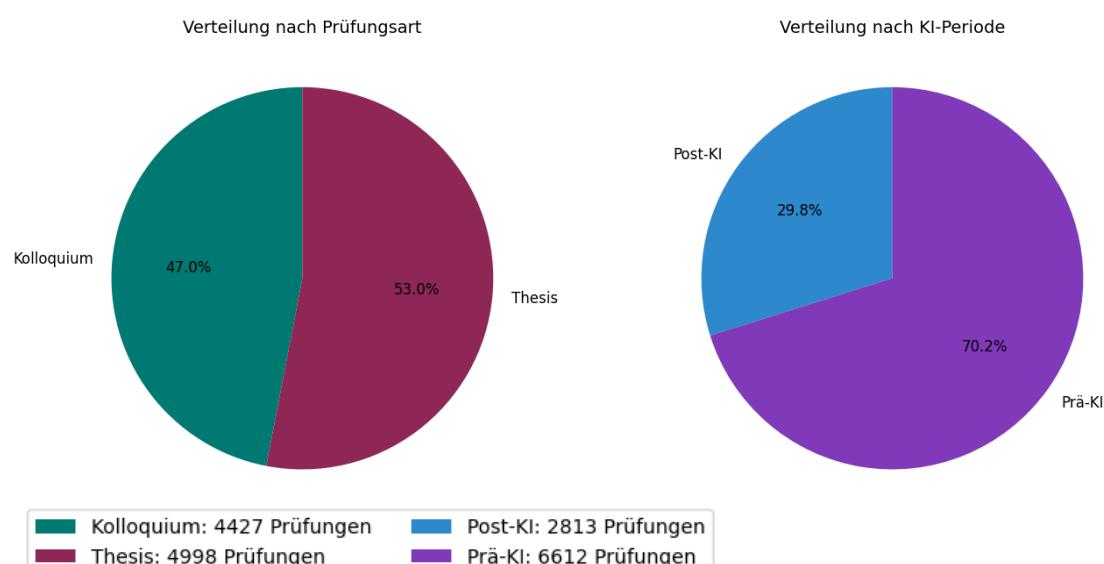


Abbildung 3.13: Verteilung der Dateneinträge nach Kategorie

### 3.3.1.6 Prüfung, ob KI Thesis-Ergebnisse verbessert

Zur Prüfung der Hypothese: „Generative KI, als Hilfsmittel für die Bearbeitung von Abschlussarbeiten, hat einen positiven Effekt auf die Ergebnisse“ wurden die folgenden prüfbaren Hypothesen aufgestellt.

- **Nullhypothese ( $H_0$ ):** Die Ergebnisse der Abschlussarbeiten vor und nach der flächendeckenden Verfügbarkeit von generative KI sind identisch.
- **Alternativhypothese ( $H_1$ ):** Die Ergebnisse der Abschlussarbeiten nach der flächendeckenden Verfügbarkeit von generative KI sind identisch.

Zur Prüfung dieser These wurde ein Mann-Whitney-U-Test durchgeführt.

Der Mann-Whitney-U-Test ergab, mit U-Wert = 2.593.615,50 und p-Wert = 0,986, keinen signifikanten Unterschied zwischen den Noten der Thesen vor und nach der Einführung generativer KI (Pett, 2016, 112ff).

Cliff's Delta quantifiziert die Effektstärke für Rangdaten und gilt bei Werten unter 0,147 als vernachlässigbar (Romano et al., 2006, S. 147; Cliff, 1993, S. 499–500). Die berechnete Effektstärke nach Cliff's Delta beträgt  $\delta = 0,064$  und liegt damit im Bereich eines sehr kleinen Effekts.

Die Ergebnisse des Mann-Whitney-U-Test zeigen keinen praktisch relevanten Einfluss durch die Einführung generativer KI auf die Notenverteilung bei Abschlussarbeiten im FB3 der Frankfurt UAS. Die Folgende Abbildung visualisiert die Ergebnisse und zeigt, dass die Kurven der Daten „Prä-KI“ und „Post-KI“ nahezu identisch sind.

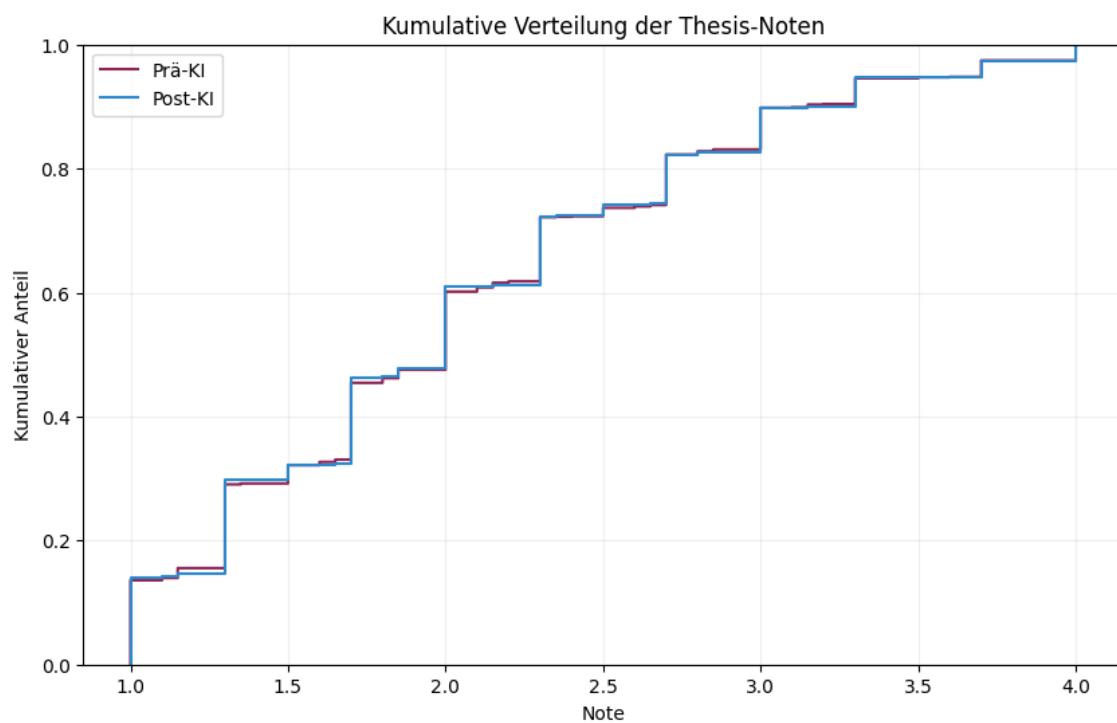


Abbildung 3.14: Kumulative Verteilung der Thesis-Noten

Ergänzend wurde eine WLS-Regression durchgeführt:

Regressionskoeffizient	0,038
P-Wert	0,298
R <sup>2</sup>	0,051

Tabelle 3.4: Ergebnisse WLS-Regression für Hypothese 4

Der Regressionskoeffizient für den KI-Effekt beträgt 0,038 bei einem p-Wert von 0,298. Damit liegt der geschätzte Einfluss deutlich unterhalb der Schwelle für statistische Signifikanz (p-Wert > 0,05), sodass die Nullhypothese ( $H_0$ ), nach der sich die Noten im Zuge der KI-Einführung nicht verändert haben, beibehalten werden muss (Wooldridge, 2019, S. 280ff). Das Modell erklärt mit einem R<sup>2</sup> von 0,051 lediglich 5,1 % der Varianz der Noten, was für sozialwissenschaftliche Untersuchungen mit aggregierten Daten im typischen Bereich liegt (Wooldridge, 2019, S. 280ff).

Insgesamt liefert die Analyse keine Evidenz dafür, dass die flächendeckende Verfügbarkeit generativer KI im betrachteten Zeitraum zu einer signifikanten Verbesserung oder Verschlechterung der Abschlussnoten geführt hat.

### 3.3.2.3 Prüfung, ob KI Kolloquien beeinflusst

Zur Prüfung der Hypothese: „Generative KI, als Hilfsmittel für die Bearbeitung von Abschlussarbeiten, mindert die Eigenleistung von Studenten und damit die Ergebnisse in mündlichen Prüfungen“ wurden die folgenden prüfbaren Hypothesen aufgestellt.

- **Nullhypothese ( $H_0$ ):** Die Ergebnisse der Kolloquien vor und nach der flächendeckenden Verfügbarkeit von generativer KI sind identisch.
- **Alternativhypothese ( $H_1$ ):** Die Ergebnisse der Abschlussarbeiten nach der flächendeckenden Verfügbarkeit von generativer KI sind identisch.

Der Mann-Whitney-U-Test ergab, mit U-Wert = 2.284.171,50 und p-Wert = 0,9998, keinen signifikanten Unterschied zwischen den Noten für Kolloquien vor und nach der KI-Einführung (Pett, 2016, S. 112).

Die Effektstärke nach Cliff's Delta ( $\delta = 0,064$ ) klassifiziert den Effekt ebenfalls als praktisch vernachlässigbar (Romano et al., 2006, S. 147; Cliff, 1993, S. 499–500). Dies deutet darauf hin, dass generative KI im betrachteten Zeitraum keinen messbaren Einfluss auf die Note der Kolloquien hatte.

Abbildung 15 visualisiert die Ergebnisse und zeigt, dass die Kurven der Daten „Prä-KI“ und „Post-KI“ erneut nahezu identisch sind.

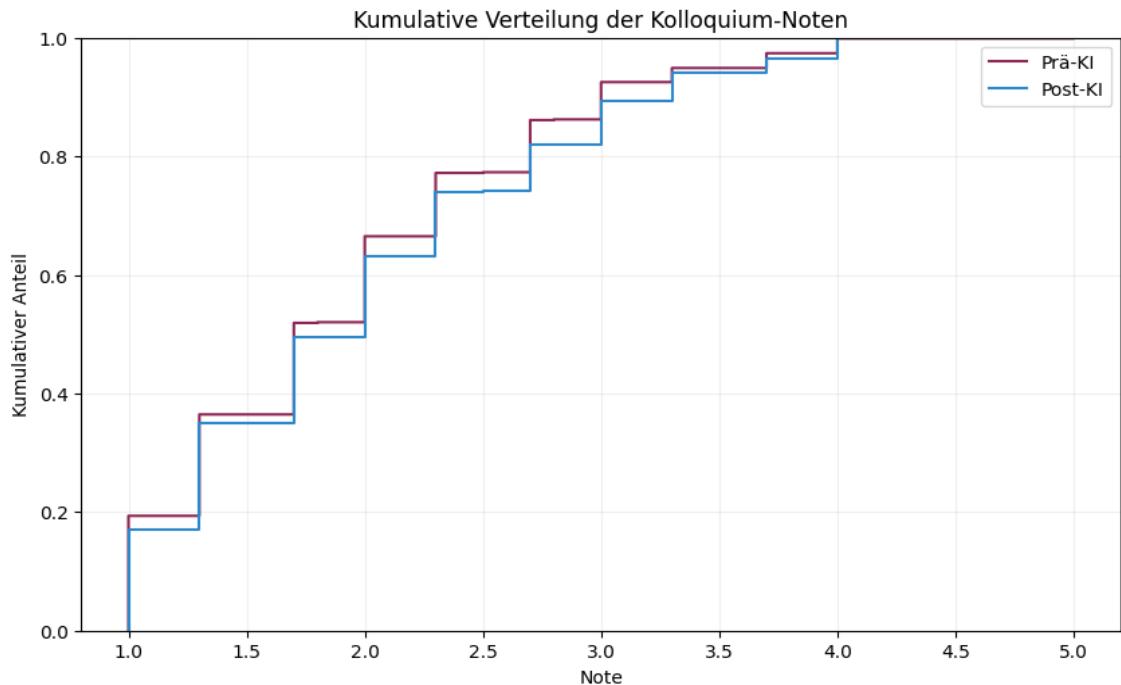


Abbildung 3.15: Kumulative Verteilung der Kolloquium-Noten

Ergänzend wurde eine WLS-Regression durchgeführt:

Regressionskoeffizient	0,065
P-Wert	0,110
R <sup>2</sup>	0,042

Tabelle 3.5: Ergebnisse WLS-Regression für Hypothese 5

Der Regressionskoeffizient für den KI-Effekt beträgt 0,065 bei einem p-Wert von 0,11. Damit liegt der geschätzte Einfluss deutlich unterhalb der Schwelle für statistische Signifikanz (p-Wert > 0,05), sodass die Nullhypothese ( $H_0$ ), nach der sich die Noten im Zuge der KI-Einführung nicht verändert haben, beibehalten werden muss (Wooldridge, 2019, S. 280ff). Das Modell erklärt mit einem R<sup>2</sup> von 0,042 lediglich 4,2 % der Varianz der Noten. (Wooldridge, 2019, S. 280ff).

Die Analyse der Kolloquien liefert ebenfalls keine Evidenz dafür, dass die flächendeckende Verfügbarkeit generativer KI im betrachteten Zeitraum zu einer signifikanten Verbesserung oder Verschlechterung der Abschlussnoten geführt hat.

### 3.4 Zwischen Wandel und Widerstand – Was die Ergebnisse zeigen

Im folgenden Kapitel werden die zentralen Ergebnisse der Hypothesenprüfungen zusammenfassend betrachtet. Dabei erfolgt eine kritische Auseinandersetzung mit den Ergebnissen sowie eine Reflexion möglicher Erklärungsansätze, Limitationen und Implikationen.

### 3.4.1 Diskussion zu den Ergebnissen der Umfrage von Lehrenden der Frankfurt UAS

Zusammenfassend zeigt sich, dass ein Großteil der Befragten negative Veränderungen in der Qualität studentischer Arbeiten seit der Verfügbarkeit generativer KI wahrnimmt. Besonders häufig wurden ein erhöhter Plagiatsverdacht und ein Rückgang des Vertrauens in die Eigenständigkeit studentischer Leistungen genannt. Viele vermuten zudem eine Abnahme der intellektuellen Leistung, die für Abschlussarbeiten erforderlich ist. Die wahrgenommene Verbesserung des Schreibstils wird dabei als Hinweis auf den verstärkten Einsatz generativer KI gewertet. Hinsichtlich der Quellenarbeit ergibt sich ein gemischtes Bild. Einige Teilnehmer sehen kaum Veränderungen, andere berichten von einer zunehmenden Nutzung fiktiver oder fehlerhafter Quellen. Dies deutet darauf hin, dass Studierende im Umgang mit KI-generierten Inhalten teils unsicher sind und Schwierigkeiten haben, die Zuverlässigkeit von Quellen angemessen zu prüfen.

Diese Ergebnisse lassen sich dahingehend auslegen, dass Lehrende sich einer ambivalenten Entwicklung gegenübersehen. Der wahrgenommene Rückgang der intellektuellen Eigenleistung sowie das gesunkene Vertrauen in die Eigenständigkeit lassen sich als zentrale Risikofaktoren interpretieren, die die wissenschaftliche Qualität von Abschlussarbeiten aus Sicht der Lehrenden gefährden. Gleichzeitig wird jedoch auch eine Verbesserung des Schreibstils festgestellt. Dieser kann als Indiz für den zunehmenden Einsatz generativer KI ausgelegt werden, jedoch symbolisiert dies auch die unterstützende Wirkung von KI. Dadurch erhalten sprachlich weniger versierte Studierende die Möglichkeit, ihre qualitativ hochwertigen Gedankengänge sprachlich zu optimieren und in einer eloquenten Form darzustellen.

### 3.4.2 Diskussion zu den Ergebnissen der historischen Daten

Die statistischen Erhebungen zur Hypothese 4 und Hypothese 5 konnten in beiden Fällen mit den unterschiedlichen Methoden des Mann-Whitney-U-Test und der WLS-Regression keine signifikanten Beweise finden. In beiden Fällen wurde die Nullhypothese angenommen, die keine signifikanten Veränderungen der Noten für Thesen oder Kolloquien erfüllt.

Die Ergebnisse der Hypothese 5 decken sich mit der Umfrage der Professoren. Diese berichteten mit großer Mehrheit keinen Effekt in Hinblick auf nachlassende Leistungen im Kolloquium seit der flächendeckenden Verfügbarkeit von generativer KI.

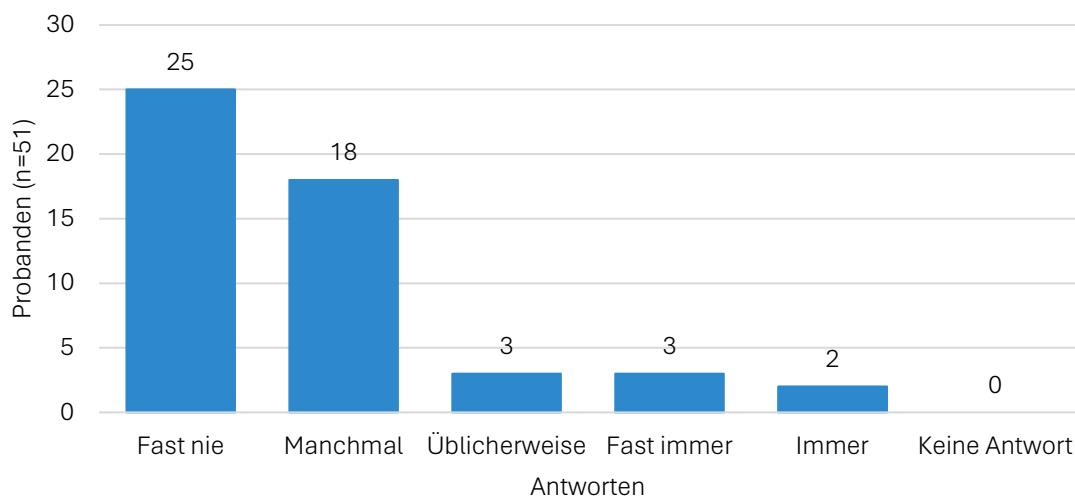


Abbildung 3.16: Umfrageergebnis - Wissensdiskrepanz zwischen Kolloquium und Abschlussarbeit

Es muss jedoch beachtet werden, dass die vorliegende Analyse mehreren methodischen Einschränkungen unterliegt, die bei der Interpretation der Ergebnisse zu berücksichtigen sind.

Die Verwendung aggregierter Notendaten anstelle individueller Prüfungsleistungen führt zu einem Verlust individueller Nutzungsmuster, wie etwa unterschiedlicher Intensitäten der KI-Nutzung von Studenten. Zudem besteht ein Ecological-Fallacy-Risiko, bei dem Gruppenmerkmale fälschlich auf individuelle Fälle übertragen werden könnten.

Weiterhin konnten wichtige Störfaktoren nicht kontrolliert werden, darunter Änderungen in Prüfungsmodalitäten, studiengangsspezifische Unterschiede in der KI-Nutzungsintensität, qualitative Aspekte der Betreuung sowie die thematische Komplexität der Arbeiten.

Messprobleme zeigen sich insbesondere in der Operationalisierung der KI-Exposition, die lediglich den Zeitpunkt der Verfügbarkeit erfasst, nicht aber die tatsächliche Nutzung.

Statistische Einschränkungen ergeben sich aus ungleichen Gruppengrößen (Prä-KI: 70,2 % vs. Post-KI: 29,8 %, siehe Kapitel 4.2.1) und der Herausforderung, kleine Effekte bei begrenzter Stichprobengröße zu detektieren.

Die nicht-signifikanten Ergebnisse können sowohl auf echte Null-Effekte hindeuten – etwa, dass KI-Kernkompetenzen wissenschaftlichen Arbeitens nicht verbessert – als auch auf methodische Artefakte zurückgeführt werden. Letztere umfassen den Verlust individueller Effekte durch Datenaggregation und den kurzen Beobachtungszeitraum von nur vier Post-KI-Semestern. Kontextfaktoren wie studiengangsspezifische Unterschiede und die langsame Adaption von KI-Tools in der Lehrpraxis könnten dominierende Einflüsse darstellen.

Für zukünftige Forschung wird ein mehrdimensionaler Ansatz empfohlen: Längsschnittstudien mit individuellen Daten, die KI-Nutzung via Lernplattform-Logs tracken und Prüfungsleistungen vor/nach KI-Einführung bei denselben Studierenden vergleichen, könnten tiefere Einsichten liefern. Ein Mixed-Methods-Design sollte qualitative und quantitative Elemente integrieren, während differenzierte Effektmessungen verschiedene KI-Nutzungsintensitäten und fachspezifische Analysen (z.B. Rechtswissenschaften vs. BWL) vergleichen.

Zusätzlich sollten erweiterte Indikatoren wie Plagiatsquoten, Literaturverzeichnisqualität und methodische Tiefe ein umfassenderes Bild der KI-Auswirkungen liefern.

Diese multidimensionale Herangehensweise könnte die komplexen Wechselwirkungen zwischen Technologieeinsatz und akademischer Leistungsbewertung präziser erfassen und somit eine fundiertere Bewertung des KI-Einflusses ermöglichen.

### **3.5 Zentrale Erkenntnisse und Implikationen**

Die vorliegende Arbeit untersuchte den Einfluss generativer KI auf Abschlussarbeiten und Kolloquien an der Frankfurt UAS. Dabei wurden sowohl die Wahrnehmung der Lehrenden als auch die Entwicklung der Abschlussnoten betrachtet. Die Umfrage zeigt, dass Lehrende generative KI vor allem als Risiko für die wissenschaftliche Qualität wahrnehmen. Besonders hervorgehoben wurden die potenzielle Abnahme der Eigenständigkeit, ein vermuteter Rückgang intellektueller Leistung sowie ein steigendes Plagiatsrisiko. Gleichzeitig wurde jedoch auch eine Verbesserung des sprachlichen Ausdrucks festgestellt, die den positiven Einfluss KI-gestützter Schreibtools verdeutlicht.

Die statistische Auswertung ergibt ein ambivalentes Bild: Trotz der geäußerten Bedenken der Lehrenden lässt sich kein signifikanter Einfluss generativer KI auf die Notenvergabe bei

Abschlussarbeiten und Kolloquien nachweisen. Die vermuteten Effekte spiegeln sich bislang nicht in den Bewertungsergebnissen wider.

Auch die Annahme, dass KI-gestützte Schreibprozesse zu Wissenslücken und Leistungsabfällen im Kolloquium führen, konnte nicht bestätigt werden. Die Ergebnisse legen vielmehr nahe, dass die Kompetenz zur mündlichen Darstellung und Reflexion weiterhin erhalten bleibt.

Die Arbeit weist jedoch Limitationen auf: Die begrenzte Stichprobengröße, der Fokus auf zwei Fachbereiche sowie die ausschließliche Erfassung des Zeitpunkts der KI-Verfügbarkeit schränken die Aussagekraft ein. Zudem ist der Beobachtungszeitraum der Post-KI-Phase für langfristige Effekte zu kurz. Zukünftige Forschung sollte individuelle KI-Nutzungsmuster detaillierter erfassen und längsschnittliche Analysen mit qualitativen Interviews kombinieren, um die tatsächliche Integration generativer KI und deren Auswirkungen besser zu verstehen. Ebenso sind hochschuldidaktische Anpassungen erforderlich, die KI-Kompetenz fördern und wissenschaftliche Integrität sichern.

### 3.6 Literaturverzeichnis

- Christ-Brendemühl, S. (2024). Leveraging Generative AI in Higher Education: An Analysis of Opportunities and Challenges Addressed in University Guidelines. European Journal Of Education. <https://doi.org/10.1111/ejed.12891>
- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin, 114(3), 494-509. (DOI: 10.1037/0033-2909.114.3.494)
- Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). Sage.
- Mayring, P. (2022). Qualitative Inhaltsanalyse: Grundlagen und Techniken (13. Aufl.).
- McKnight, P. E., & Najab, J. (2010). Mann-Whitney U Test. The Corsini Encyclopedia of Psychology. doi: 10.1002/9780470479216.corpsy0524
- OpenAI (2022) Introducing ChatGPT. <https://openai.com/index/chatgpt/>. [Abgerufen am 22.06.2025]
- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London Edinburgh And Dublin Philosophical Magazine And Journal Of Science, 50(302), 157–175. <https://doi.org/10.1080/14786440009463897>
- Pett, M. A. (2016). Nonparametric statistics for health care research. Sage Publications.
- Romano, J., Kromrey, J. D., Coraggio, J., Skowronek, J., & Devine, L. (2006). Exploring methods for evaluating group differences on the NSSE and other surveys: Are the t-test and Cohen's d indices the most appropriate choices? Annual Meeting of the Southern Association for Institutional Research, S. 147.
- Wooldridge, J. M. (2019). Introductory Econometrics: A Modern Approach (7th ed.). Cengage Learning.

# **Generative KI bei Abschlussarbeiten aus Sicht der Studierenden**

## 4 Eigenleistung bei wissenschaftlichen Arbeiten im Umbruch

### *Nachweisbarkeit von Eigenleistung im wissenschaftlichen Arbeiten mit generativer KI*

Anuja Ariyara

Zakaria Salah

#### **Abstract**

Die zunehmende Nutzung generativer KI im Studium stellt das Verständnis wissenschaftlicher Eigenleistung vor neue Herausforderungen. Diese Arbeit untersucht, wie Studierende Eigenleistung im Zeitalter generativer KI definieren und welche Maßnahmen sie als glaubwürdig für den Nachweis ansehen. Grundlage ist eine quantitative Online-Befragung unter Studierenden des Bachelorstudiengangs Wirtschaftsinformatik (N = 66). Geprüft werden drei Hypothesen zum Zusammenhang von KI-Nutzung, Selbstwirksamkeit, Regelwissen und sozialem Rückhalt mit Eigenleistungsverständnis und Offenlegungsbereitschaft.

Die Ergebnisse zeigen, dass Studierende Eigenleistung nicht über Vermeidung, sondern über reflektierte Nutzung definieren. Selbstwirksamkeit wirkt nur dann signifikant, wenn keine positive Haltung zu KI besteht. Regelkenntnis und Rückhalt haben kaum Einfluss auf Offenlegung.

Die Ergebnisse unterstreichen die Notwendigkeit kontextbezogener Nachweismethoden und einer didaktischen Weiterentwicklung des Leistungsbegriffs.

**Schlagwörter:** Generative KI, Eigenleistung, Nachweisbarkeit, Hochschullehre, Transparenz, Offenlegung, Reflexion, Selbstwirksamkeit, akademische Integrität.

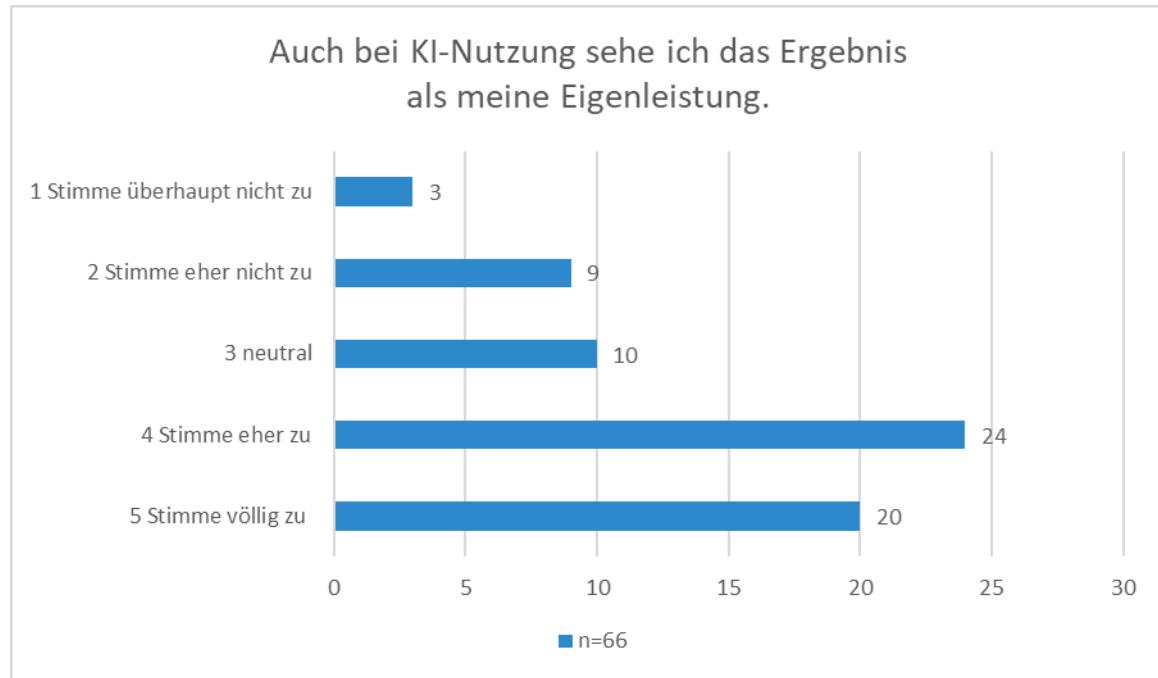


Tabelle 4.1: Bewertung der Eigenleistung bei Nutzung von KI

## 4.1 Zwischen Anspruch und Anpassung

„Die [...] Welt löst nicht nur Probleme, indem Wissenschaft und Technik Probleme lösen; sie schafft auch Probleme“ – so der deutsche Philosoph Jürgen Mittelstraß. Für ihn gestalten Wissenschaft und Technik nicht nur unsere Umwelt, sondern prägen zunehmend auch Denken und Handeln, bis hin zur Abhängigkeit von den eigenen Erfindungen (Mittelstraß, 1990, S. 149).

Gerade im Hochschulkontext wird diese Dynamik sichtbar. Was als technologische Innovation beginnt, wird schnell Teil alltäglicher Lern- und Arbeitsprozesse. Eine Längsschnittstudie zeigt, dass 2025 bereits 91,6 % der Studierenden in Deutschland Tools generativer künstlicher Intelligenz (im Folgenden KI) wie ChatGPT, Microsoft Copilot oder DeepL nutzen, vor allem zur Unterstützung wissenschaftlicher Arbeiten (Garrel & Mayer, 2025, S.19, 26). Der technologische Wandel fordert nicht nur neue Lernformate, sondern stellt auch den Wert individueller Leistung neu zur Debatte.

Die Integration generativer KI verlangt von Studierenden, ihre wissenschaftliche Arbeitsweise an veränderte Bedingungen anzupassen. Dabei geht es nicht nur um die Zulässigkeit der Nutzung, sondern um die Frage, wie Eigenverantwortung, Transparenz und reflektiertes Handeln unter neuen Rahmenbedingungen noch glaubwürdig belegt werden können. Zwischen institutionellen Vorgaben, persönlichen Haltungen und sozialen Erwartungen entsteht ein Spannungsfeld, das eine Neubewertung des Eigenleistungsbegriffs erfordert.

Diese Untersuchung ist Teil eines kursübergreifenden empirischen Forschungsprojekts zur Nutzung generativer KI im Studium. Im Fokus steht die Forschungsfrage: Wie definieren Studierende Eigenleistung im Zeitalter generativer KI und wie stellen sie sicher, diese bei wissenschaftlichen Arbeiten nachweisen zu können? Die empirische Basis bildet eine quantitative Online-Befragung unter Studierenden des Bachelorstudiengangs Wirtschaftsinformatik an der Frankfurt University of Applied Sciences.

Daraus ergeben sich drei Hypothesen:

1. Wenn Studierende KI-Tools nutzen, dann verstehen sie Eigenleistung als reflektierten und transparenten Umgang damit und nicht als ihre vollständige Vermeidung.
2. Wenn Studierende eine höhere Selbstwirksamkeit aufweisen, dann definieren sie Eigenleistung strenger, unabhängig davon, ob sie KI-generierte Inhalte als Eigenleistung ansehen.
3. Wenn Studierende sich gut über institutionelle Vorgaben zum KI-Einsatz informiert fühlen oder sozialen Rückhalt durch Lehrende und Mitstudierende erfahren, zeigen sie höhere Bereitschaft, ihre Eigenleistung transparent von KI-generierten Inhalten abzugrenzen.

Ziel ist es, das Verständnis von Eigenleistung aus studentischer Perspektive in Zeiten generativer KI nachzuvollziehen und einzuordnen. Die Arbeit zeigt, wie KI-Tools nicht nur die wissenschaftliche Arbeitsweise beeinflussen, sondern auch den Anspruch an Selbstständigkeit, Verantwortung und Transparenz verändern. Der Fokus liegt auf schriftlichen Prüfungsleistungen, da sie typischerweise außerhalb von Prüfungsaufsicht entstehen und daher einen besonders hohen Anspruch an nachweisbare Eigenleistung stellen.

Die Arbeit gliedert sich wie folgt: Zunächst wird im theoretischen Rahmen der Begriff der Eigenleistung definiert sowie psychologische und soziale Einflussfaktoren betrachtet, darunter auch die Bedeutung der Selbstwirksamkeit, die im weiteren Verlauf näher erläutert wird. Anschließend wird der Forschungsstand zur KI-Nutzung und zur Wahrnehmung von Eigenleistung im Studium

analysiert. Es folgen das methodische Vorgehen, die Darstellung und Auswertung der Ergebnisse entlang der Hypothesen sowie eine abschließende Diskussion mit Fazit und Implikationen.

## 4.2 Grundlagen zu Eigenleistung und KI im wissenschaftlichen Arbeiten

Der theoretische Rahmen dieser Arbeit behandelt drei zentrale Bereiche, nämlich die begriffliche und kontextuelle Einordnung von Eigenleistung, psychologische und soziale Einflussfaktoren sowie den Forschungsstand zur Nutzung generativer KI und zur studentischen Wahrnehmung von Eigenleistung.

### 4.2.1 Definition Eigenleistung

Im Duden wird Eigenleistung als eine „aus eigener Kraft erbrachte Leistung“ beschrieben (Duden.de, 2024). Für den akademischen Kontext muss dieser Begriff differenzierter gefasst werden. Hochschuldidaktisch wird Eigenleistung als Ergebnis eines aktiven und kognitiv fordernden Lernprozesses verstanden, bei dem Studierende Inhalte selbstständig durchdringen, reflektieren und bewerten. Studierendenzentrierte Lehre zielt darauf ab, selbstverantwortliches Lernen zu fördern, bei dem individuelle Urteilsbildung und persönliche Auseinandersetzung mit dem Fachgegenstand zentral sind (Ulrich und Brieden, 2021, S. 4; Cromme, 2021, S. 65f.).

Mit dem Aufkommen generativer KI haben sich die Anforderungen an Eigenleistung grundlegend verändert. Generative KI-Tools wie ChatGPT können Texte erzeugen, die den äußeren Anforderungen wissenschaftlicher Arbeiten genügen, ohne dass zwingend ein eigener gedanklicher Beitrag vorliegt. Es ist festzuhalten, dass vollständig durch KI-Tools generierte Texte nicht als eigenständige Prüfungsleistung anerkannt werden können. Die Nutzung solcher Tools ist zulässig, muss aber transparent gemacht und inhaltlich klar abgegrenzt werden, um Täuschungsvorwürfe zu vermeiden (Fleischmann, 2025, S. 237). Die Verantwortung für die Qualität und die inhaltliche Richtigkeit der Prüfungsleistung verbleibt bei den Studierenden, wenn KI-basierte Hilfsmittel eingesetzt wurden (ebd.).

Die Verantwortung für die Ausgestaltung von Prüfungsleistungen liegt gemäß § 50 Abs. 1 Nr. 1 HessHG bei den Hochschulen. Die Frankfurt University of Applied Sciences regelt in § 24 ihrer Allgemeinen Bestimmungen, dass der Einsatz generativer KI zulässig ist, sofern Studierende aktiv steuernd eingreifen, den eigenen gestalterischen Anteil sicherstellen und die Nutzung transparent machen. Bei Verstößen kann die Leistung als Täuschungsversuch gewertet und mit „nicht bestanden“ bewertet werden (Frankfurt University of Applied Sciences, 2023a, § 24; Gumprecht et al., 2024, S. 6). Zusätzlich ist mit jeder schriftlichen Arbeit eine unterschriebene Eigenständigkeitserklärung abzugeben, in der die selbstständige Bearbeitung versichert und der KI-Einsatz offen deklariert wird (Frankfurt University of Applied Sciences, 2023b). Zudem bleibt festzuhalten, dass KI-Tools keine Urheber im rechtlichen Sinn darstellen können, da ihnen die Fähigkeit zur persönlichen geistigen Schöpfung fehlt (Salden und Leschke, 2024, S. 16). Urheberschaft ist gemäß § 2 Abs. 2 UrhG an menschliches schöpferisches Handeln gebunden und kann nur dann den Studierenden zugesprochen werden, wenn sie wesentliche eigene Beiträge leisten und die KI lediglich unterstützend einsetzen (ebd.).

In der Praxis entsteht dadurch die Herausforderung, dass die Eigenleistung schwer überprüfbar wird, wenn große Teile einer Arbeit durch KI erzeugt werden. Deshalb wird empfohlen, Aufgaben so zu gestalten, dass sie kreative, individuelle und komplexe Lösungsansätze erfordern (Hof, 2025, S. 633).

Dies kann durch neuartige Fragestellungen, Reflexionsaufgaben oder mündliche Erläuterungen erreicht werden. Ergänzend dazu sollen Studierende zum reflektierten Umgang mit generativer KI befähigt werden, ein generelles Verbot erscheint aufgrund fehlender Kontrollmöglichkeiten nicht zielführend (ebd.).

Insgesamt zeigt sich, dass sich das Verständnis von Eigenleistung im Kontext generativer KI erweitert hat. Neben der traditionellen Anforderung der Selbstständigkeit rücken Transparenz, Verantwortung und der reflektierte Einsatz technischer Hilfsmittel in den Vordergrund. Eigenleistung besteht im Zeitalter generativer KI darin, den Einsatz generativer KI kritisch zu reflektieren, Ergebnisse gezielt weiterzuentwickeln und diese nachvollziehbar in den eigenen wissenschaftlichen Arbeitsprozess zu integrieren.

#### 4.2.2 Psychologische und soziale Einflussfaktoren der KI in der Hochschullehre

Zwischen individueller Kompetenz und digitaler Unterstützung entsteht durch KI-Nutzung im Studium ein neues Spannungsfeld. Theoretisch lässt sich dieses Phänomen unter Rückgriff auf psychologische, technikbezogene und bildungssoziologische Modelle fassen.

Ein zentrales psychologisches Konzept im Zusammenhang mit Eigenverantwortung und Leistungserbringung ist die Selbstwirksamkeit nach Bandura (1977). Sie beschreibt die Überzeugung eines Individuums, „*dass es in der Lage ist, bestimmte Handlungen erfolgreich auszuführen, um gewünschte Ergebnisse zu erzielen*“ (Bandura, 1977, S. 193). Selbstwirksamkeit beeinflusst nicht nur das Verhalten, sondern auch die Zielverfolgung und den Umgang mit Herausforderungen. Auf das Thema der akademischen Eigenleistung übertragen bedeutet dies: Studierende mit einem hohen Maß an Selbstwirksamkeit könnten weniger auf KI-Hilfen angewiesen sein oder reflektierter mit deren Einsatz umgehen, da sie Vertrauen in ihre eigenen Kompetenzen haben.

Auf systemischer Ebene ist im Kontext des Unified Theory of Acceptance and Use of Technology Modells (im Folgenden UTAUT-Modell; Venkatesh et al., 2003) insbesondere der Faktor „Social Influence“ (soziale Einflussnahme) bedeutsam. „Social Influence“ beschreibt den Grad, zu dem Individuen wahrnehmen, dass wichtige Bezugspersonen, wie Lehrende oder Kommiliton\*innen, erwarten, dass sie bestimmte Technologien nutzen (Venkatesh et al., 2003, S. 451). Empirische Studien zeigen, dass insbesondere im Hochschulkontext der soziale Einfluss durch Lehrende und Mitstudierende einen erheblichen Einfluss auf die Nutzung von KI-Anwendungen hat. So fanden Faraon et al., dass soziale Einflussfaktoren, insbesondere durch Lehrende, einen maßgeblichen Einfluss auf die Nutzungsintention von KI-Anwendungen bei Studierenden haben (Faraon et al., 2025, S. 4). Dies unterstreicht die Notwendigkeit, den Einsatz generativer KI nicht nur individuell, sondern auch institutionell und sozial zu gestalten, eine Forderung, die auch Fleischmann (2023) erhebt. Die Integration solcher Technologien sollte daher nicht nur technisch, sondern auch sozial und institutionell begleitet werden, um einen verantwortungsvollen Einsatz zu gewährleisten.

Die zuvor beschriebenen psychologischen und sozialen Einflussfaktoren werden auch durch empirische Studien im Hochschulbereich gestützt. Zawacki-Richter et al. (2019) zeigen in ihrem systematischen Review, dass KI-Anwendungen im Hochschulbereich vor allem in den Bereichen „Assessment and Evaluation“, „Profiling and Prediction“, „Adaptive Systems and Personalisation“ sowie „Intelligent Tutoring Systems“ eingesetzt werden (Zawacki-Richter et al., 2019, S. 11–18). Im Bereich „Assessment and Evaluation“ werden insbesondere automatisierte Bewertungssysteme und Feedback-Mechanismen hervorgehoben. Die Autoren diskutieren ethische Herausforderungen und Risiken, wie Datenschutz und Privatsphäre, die mit der Nutzung von KI in der Bildung verbunden sind, und betonen die Notwendigkeit klarer institutioneller Regelungen, um diese Herausforderungen zu adressieren (Zawacki-Richter et al., 2019, S. 20–22).

Auch wenn Zawacki-Richter et al. (2019) die konkrete Nutzung von KI durch Studierende nicht detailliert behandeln, lässt sich aus ihrer Kritik am Mangel institutioneller, ethischer und pädagogischer Steuerung ableiten, dass der KI-Einsatz im Hochschulbereich teils in einer regulativen Grauzone erfolgt, was mutmaßlich auch die informelle Nutzung durch Studierende betrifft. Diese mangelnde Orientierung ist ein zentraler Aspekt bei der Betrachtung von Eigenleistung: Ohne klare Rahmenbedingungen bleibt den Studierenden die Entscheidung über den „richtigen“ Umgang mit KI weitgehend selbst überlassen.

Genau hier setzt Fleischmann (2023) an. In seiner Analyse zur Rolle von ChatGPT in der Hochschullehre beschreibt er, dass die Nutzung von künstlicher Intelligenz die Hochschullehre vor neue Herausforderungen stellt. Insbesondere die Frage der Eigenleistung wird durch die Integration generativer KI komplexer. Der Autor hebt hervor, dass Lehrende und Studierende lernen müssen, die Möglichkeiten und Grenzen solcher Tools zu verstehen und reflektiert damit umzugehen. Fleischmann plädiert für eine bewusste Anpassung der Curricula und Prüfungsformate, um den Einsatz von KI regelgeleitet und diskursiv zu gestalten (Fleischmann, 2023, S. 22).

Die Forderung eines diskursiven und regelgeleiteten Umgangs mit generativer KI ist daher eine Forderung, die wieder auf der theoretischen Ebene in den Begriff der „Social Influence“ aus dem UTAUT-Modell verweist. Kurzum lässt sich festhalten, dass das Verhalten anderer (z. B. Lehrende, Kommiliton\*innen) und institutionelle Vorgaben einen erheblichen Einfluss auf die individuelle Technologienutzung haben (Venkatesh et al., 2003, S. 451).

#### **4.2.3 Eigenleistung und KI-Nutzung im Fokus aktueller empirischer Forschung**

Der zunehmende Einsatz generativer KI-Tools im Hochschulkontext hat nicht nur die Anforderungen an Eigenleistung in wissenschaftlichen Arbeiten verändert, sondern auch zahlreiche empirische Studien hervorgebracht, die sich mit deren Nutzung, Wahrnehmung und Bewertung durch Studierende befassen. Während sich die Nutzung zunehmend, als Bestandteil akademischer Routinen etabliert hat (Garrel & Mayer, 2025, S. 2–3), werden Fragen zur Abgrenzung eigener Leistung, zum institutionellen Rahmen und zu psychologischen Einflussfaktoren bislang nur vereinzelt thematisiert. Im Hinblick auf die eigene Erhebung, die sich auf Studierende der Wirtschaftsinformatik an der Frankfurt University of Applied Sciences konzentriert, werden im Folgenden vor allem Befunde berücksichtigt, die wirtschaftswissenschaftliche Studiengänge betreffen. Der folgende Abschnitt bietet einen Überblick über zentrale empirische Studien, die für das Thema dieser Arbeit relevant sind.

Die quantitative Längsschnittstudie von Garrel und Mayer (2025) verdeutlicht die starke Verbreitung generativer KI an deutschen Hochschulen. Während 2023 noch 63,2 % der befragten Studierenden KI-Tools nutzten, lag dieser Anteil 2025 bei 91,6 % (Garrel & Mayer, 2025, S. 19). Besonders hoch ist die Nutzung in den Ingenieurwissenschaften mit 95,5 %, gefolgt von Natur-, Gesundheits-, Wirtschafts- und Sozialwissenschaften mit jeweils über 90 %. Am häufigsten werden ChatGPT (kostenfrei und kostenpflichtig), DeepL, Microsoft Copilot und DALL-E genannt, wobei ChatGPT dominiert (Garrel & Mayer, 2025, S. 2, 26). In den Wirtschaftswissenschaften werden KI-Tools insbesondere zur Klärung von Verständnisfragen, zum wissenschaftlichen Schreiben (Textverarbeitung, -analyse und -erstellung), zur Recherche sowie zur Problemlösung eingesetzt (Garrel & Mayer, 2025, S. 27–33). Der Umgang mit KI ist an deutschen Hochschulen uneinheitlich. Hochschulen integrieren tendenziell praxisnahe Formate während Universitäten forschungsorientierte Ansätze bevorzugen. Während einige Einrichtungen auf Reflexion und offene Formate setzen, verfolgen andere klare Vorgaben oder entwickeln eigene Anwendungsszenarien. Diese Unterschiede können zu Unsicherheiten im studentischen Umgang mit KI beitragen (Garrel & Mayer, 2025, S. 13).

Diese Unsicherheit zeigt sich auch in der Studie von Gottschling, Seidl und Vonhof (2024). Nur 13,6 % der Studierenden deklarieren ihren KI-Einsatz bei wissenschaftlichen Arbeiten vollständig, obwohl viele die Tools gezielt und kompetent einsetzen (Gottschling et al., 2024, S. 130). Die Nutzung erfolgt oft im Graubereich, da es an klaren institutionellen Regelungen fehlt (Gottschling et al., 2024, S. 133). Zugleich wünschen sich 75 % der Befragten mehr Orientierung im Umgang mit KI, was auf einen deutlichen Bedarf an institutioneller Begleitung verweist (Gottschling et al., 2024, S. 131).

Darüber hinaus gewinnen psychosoziale Einflussfaktoren an Bedeutung. Die Untersuchung von Sallam et al. (2024) zeigt, dass die Nutzung von ChatGPT durch wahrgenommene Nützlichkeit ( $B = 0.398$ ), kognitive Faktoren wie Vertrauen in die eigene Kompetenz ( $B = 0.276$ ) und sozialen Einfluss ( $B = 0.413$ ) beeinflusst wird (Sallam et al., 2024, S. 8–10). Kognitive Faktoren umfassen Aspekte der Selbstwirksamkeit, also die Überzeugung, KI zielgerichtet nutzen zu können. Soziale Einflussgrößen beziehen sich auf das Umfeld, etwa ob Lehrende, Mitstudierende oder Institutionen den KI-Einsatz unterstützen. Zusammengenommen erklären diese Faktoren einen Großteil der Nutzung und Einstellung gegenüber ChatGPT. Die Ergebnisse stützen Hypothese 2 und 3, die Selbstwirksamkeit und sozialen Rückhalt im reflektierten KI-Einsatz untersuchen.

Eine internationale Ergänzung liefert die Studie von Almassaad et al. (2024) an der King Saud University. Auch dort zeigt sich eine breite Nutzung generativer KI-Tools, insbesondere zur Ideenfindung, Textüberarbeitung und Übersetzung. In vielen Hochschulsystmen bestehen Unklarheiten zu verbindlichen Regelungen. Nur 19,2 % der Befragten kannten konkrete Vorgaben. Die Studierenden bewerten KI meist positiv, etwa wegen Zeitersparnis und einfacher Bedienbarkeit, äußern aber auch Bedenken bezüglich Plagiatsrisiken und inhaltlicher Zuverlässigkeit (Almassaad et al., 2024, S. 9). Die Ergebnisse betonen die Bedeutung eines reflektierten und kontextsensiblen Umgangs mit KI für den Nachweis von Eigenleistung.

Der aktuelle Forschungsstand bestätigt, dass generative KI längst im universitären Alltag angekommen ist. Gleichzeitig zeigen die Studien, dass Fragen der Eigenleistung, Transparenz und Verantwortung häufig unzureichend behandelt werden. Bisherige Untersuchungen fokussieren vor allem auf Nutzungs frequenz und allgemeine Einstellungen, ohne systematisch zu analysieren, wie Studierende Eigenleistung im Kontext von KI definieren und nachweisen. Auch psychosoziale Faktoren wie Selbstwirksamkeit oder sozialer Rückhalt werden bislang nur punktuell berücksichtigt. Es fehlt an theoriebasierten Studien, die diese Aspekte, etwa auf Basis von Bandura (1977) oder Venkatesh et al. (2003) empirisch verknüpfen.

Die vorliegende Arbeit greift diese Lücke auf. Sie untersucht nicht nur, ob und wie Studierende KI nutzen, sondern auch, wie sie ihren wissenschaftlichen Beitrag im Zusammenspiel mit KI definieren, reflektieren und abgrenzen und leistet damit einen Beitrag zur empirisch fundierten Diskussion über akademische Eigenleistung im Zeitalter generativer KI.

## 4.3 Methodisches Vorgehen

Das methodische Vorgehen wurde bereits im übergeordneten Kapitel 1.2 Datenerhebung und Stichprobe dargestellt. Aufbauend auf dem gemeinsamen Forschungsdesign und der identischen Datenbasis wird im Folgenden lediglich die konkreten Schritte der Datenaufbereitung, Kodierung und Analyseverfahren dargestellt, die für die vorliegende Auswertung angewendet wurden.

### 4.3.1 Datenanalyse

Zur Auswertung der erhobenen Daten wurden quantitative statistische Verfahren eingesetzt. Die Analysen erfolgten unter Einsatz von deskriptiver Statistik, um zentrale Tendenzen (Mittelwerte,

Standardabweichungen) sowie Häufigkeitsverteilungen zu beschreiben. Darüber hinaus kamen inferenzstatistische Verfahren zum Einsatz, die sich an den Hypothesen sowie den soziodemografischen Merkmalen der Stichprobe orientieren.

#### 4.3.2 Kodierung und Datenaufbereitung

##### Soziodemografische Variablen

Diese Variablen beschreiben persönliche Grundmerkmale der Teilnehmenden:

Variable	Bezeichnung (im Datensatz)	Kodierung
Alter	Alter_neu	19 = 1, 20 = 2, 21 = 3, 22 = 4, 23 = 5, 24 = 6, 25 = 7, 26 = 8, 27 = 9, 28 = 10, über 30 = 11
Geschlecht	Geschlecht_neu	Divers = 1, Männlich = 2, Weiblich = 3 (Für inferenzstatistische Analysen wurde „Divers“ ausgeschlossen, da nur einmal vertreten.)
Semesteranzahl	Semester_neu	2. Semester = 1, 4. Semester = 2, 6. Semester = 3, über 8. Semester = 4

Tabelle 4.2: Kodierung soziodemografische Variablen

## Themenspezifische Kontextvariablen

Diese Variablen beziehen sich auf das Verhalten, die Erfahrung oder Einstellungen der Studierenden im Zusammenhang mit generativer KI:

Variable	Bezeichnung (im Datensatz)	Kodierung
Hausarbeiten pro Jahr	Hausarbeiten_Anzahl_neu	1 = 1, 2 = 2, 3 = 3, 4 = 4, 5 = 5, über 8 = 6
Erfahrung mit ChatGPT	Erfahrung_ChatGPT	1 = sehr gering bis 7 = sehr hoch
Nutzung von ChatGPT im Studium	Nutzung_ChatGPT_neu	1 = Nie, 2 = Selten, 3 = Gelegentlich, 4 = Häufig, 5 = Bei jeder wissenschaftlichen Arbeit
Offenlegung bei Vorbildwirkung	Offenlegung_wenn_andere	1 = sehr unwahrscheinlich bis 5 = sehr wahrscheinlich
Informationsstand zu KI-Regelungen	Info_KI_Regeln	1 = überhaupt nicht bis 5 = sehr gut
* Einstellungs- & Bewertungsvariablen	-	Einheitliche 5er-Likert-Skala: 1 = stimme überhaupt nicht zu 2 = stimme eher nicht zu 3 = neutral 4 = stimme eher zu 5 = stimme völlig zu

Tabelle 4.3: Kodierung themenspezifische Kontextvariablen<sup>9</sup>

### 4.3.3 Statistische Verfahren

Für die Analyse der erhobenen Daten wurde ein zweistufiges statistisches Vorgehen gewählt, das sowohl beschreibende als auch prüfende Verfahren umfasst. Zunächst erfolgte eine explorative Darstellung der Daten, um zentrale Kennwerte wie Mittelwerte, Streuungen und Verteilungen sichtbar zu machen. Aufbauend auf diesen Befunden wurden weiterführende inferenzstatistische Methoden eingesetzt, um gezielte Zusammenhänge zwischen Variablen zu analysieren und Gruppenvergleiche vorzunehmen. Die Auswahl der Verfahren, darunter Korrelationsanalysen nach Pearson, t-Tests für unabhängige Stichproben und lineare Regressionsmodelle, orientierte sich am Skalenniveau der erhobenen Merkmale sowie an den hypothesengeleiteten Fragestellungen. Die konkrete Anwendung dieser Verfahren und die zugehörigen Ergebnisse werden im folgenden Kapitel im Detail dargestellt.

## 4.4 Eigenleistung unter dem Mikroskop

In diesem Kapitel werden die Ergebnisse der quantitativen Untersuchung vorgestellt. Die Auswertung gliedert sich in zwei Abschnitte: Zunächst werden hypothesesgeleitete Analysen unter Einsatz

---

9 Diese Kodierung wurde auf alle Items angewendet, die sich auf Eigenleistung, KI-Nutzung, Selbstwirksamkeit und Nachweismethoden beziehen.

verschiedener statistischer Verfahren. Anschließend werden weitere Ergebnisse zu Zusammenhängen zwischen Einstellungen, Erfahrungen und dem Verständnis von Eigenleistung im Umgang mit generativer KI dargestellt, entlang ergänzender Gruppenvergleiche und Regressionsmodelle.

Zum Schluss werden die Befunde zu bevorzugten Nachweismethoden von Eigenleistung im KI-Kontext zusammengefasst.

#### 4.4.1 Zusammenhang zwischen KI-Nutzung und kritischer Reflexion

Hypothese 1: Wenn Studierende KI-Tools nutzen, dann verstehen sie Eigenleistung als reflektierten und transparenten Umgang damit und nicht als ihre vollständige Vermeidung.

Zur Überprüfung dieser Hypothese wurde eine Pearson-Korrelation zwischen der Nutzung von ChatGPT und der kritischen Reflexion von KI-generierten Inhalten (Reflexion\_Kritik) berechnet. Das Ergebnis zeigte einen signifikanten positiven Zusammenhang ( $r = 0,253$ ,  $p = 0,040$ ,  $n = 66$ ).

		Nutzung_ChatGPT	Reflexion_Kritik
Nutzung_ChatGPT	Pearson-Korrelation	1	0,253*
	Sig. (2-seitig)		0,040
	N	66	66
Reflexion_Kritik	Pearson-Korrelation	0,253*	1
	Sig. (2-seitig)	0,040	
	N	66	66

Tabelle 4.4: Korrelation zwischen Nutzung\_ChatGPT und Reflexion\_Kritik<sup>10</sup>

Der signifikante positive Zusammenhang legt nahe, dass Studierende, die KI-Tools häufiger nutzen, ihre Eigenleistung eher im Sinne eines reflektierten und transparenten Umgangs mit KI verstehen, nicht im Sinne vollständiger Vermeidung.

#### 4.4.2 Zusammenhang zwischen Selbstwirksamkeit und Eigenleistungsverständnis

Hypothese 2: Wenn Studierende eine höhere Selbstwirksamkeit aufweisen, dann definieren sie Eigenleistung strenger, unabhängig davon, ob sie KI-generierte Inhalte als Eigenleistung ansehen.

---

10 Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

#### **4.4.2.1 Konstruktion und Reliabilitätsüberprüfung der Eigenleistung\_Skala und weitere Vorgehensweise**

Zur Messung eines anspruchsvollen Eigenleistungsverständnisses wurde eine summative Skala gebildet („Eigenleistung\_Skala“), die folgende fünf Items kombiniert:

- Fähigkeit zur Korrektur und Unterscheidung von KI-generierten Inhalten
- Entwicklung eigener Themen oder Fragestellungen
- Verzicht auf Copy-Paste aus KI-Antworten
- Kreativer Umgang mit Prompts
- Reflexion und Kritik an den generierten Inhalten

Die Skala spiegelt ein aktives, durchdachtes Verständnis von Eigenleistung im Zeitalter generativer KI wider.

Die Reliabilitätsanalyse ergab ein Cronbach's Alpha von 0,596, was für explorative Studien als akzeptabel gilt. Andere potenzielle Variablen, etwa eigensprachliche Formulierung mit oder ohne KI, wurden bewusst ausgeschlossen, da mehrere vorherige Testungen zeigten, dass deren Integration die interne Konsistenz der Skala senkte (Alpha-Werte < 0,55). Aus wissenschaftlicher Sicht ist es bei geringer Itemanzahl wichtiger, eine kohärente Konstruktrepräsentation mit moderater Reliabilität zu erzielen, als durch thematisch abweichende Items eine rechnerisch größere Skala zu konstruieren, deren interne Stimmigkeit nicht gegeben ist.

Anschließend wurden zwei lineare Regressionsanalysen durchgeführt: In der ersten Analyse diente *Selbstwirksamkeit\_ohne\_KI* als alleiniger Prädiktor für die *Eigenleistung\_Skala*. In der zweiten Analyse wurde zusätzlich die Variable *KI\_Eigenleistung* als Kontrollvariable einbezogen, um deren potenziellen Einfluss auf die abhängige Variable statistisch zu kontrollieren.

#### **4.4.2.2 Regressionsergebnisse**

Die einfache lineare Regression mit *Selbstwirksamkeit\_ohne\_KI* als Prädiktor ergab einen signifikanten positiven Effekt ( $\beta = 0,248$ ,  $p = 0,045$ ), wobei 6,1 % der Varianz in der *Eigenleistung\_Skala* erklärt wurde. Die Ergebnisse deuten darauf hin, dass Studierende mit höherer Selbstwirksamkeit dazu neigen, striktere Kriterien an Eigenleistung anzulegen.

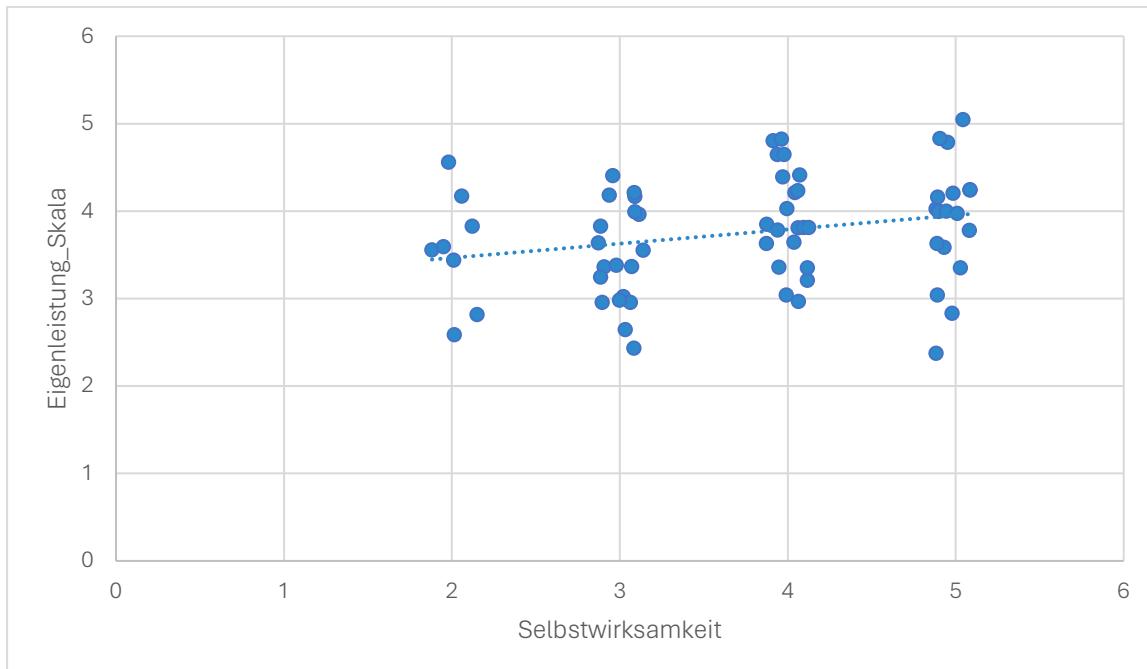


Abbildung 4.1: Zusammenhang zwischen Selbstwirksamkeit und Eigenleistungsverständnis

Wie in Abbildung 1 ersichtlich, zeigt sich ein leichter positiver Trend: Studierende mit höherer Selbstwirksamkeit (x-Achse) erzielen im Mittel höhere Werte auf der Eigenleistung\_Skala (y-Achse).

Im erweiterten Modell wurde zusätzlich die Kontrollvariable „KI\_Eigenleistung“ einbezogen, d. h. die Zustimmung zur Aussage, dass auch KI-unterstützte Ergebnisse eigene Leistung sein können. Unter dieser Kontrolle verlor die Selbstwirksamkeit an statistischer Signifikanz ( $\beta = 0,210$ ;  $p = 0,094$ ), der Effekt blieb jedoch inhaltlich bestehen.

#### 4.4.3 Zusammenhang zwischen Regelwissen, sozialem Rückhalt und KI-Abgrenzung

Hypothese 3: Wenn Studierende sich gut über institutionelle Vorgaben zum KI-Einsatz informiert fühlen oder sozialen Rückhalt durch Lehrende und Mitstudierende erfahren, zeigen sie höhere Bereitschaft, ihre Eigenleistung transparent von KI-generierten Inhalten abzugrenzen.

##### 4.4.3.1 Konstruktion und Reliabilität der Offenlegung\_Skala und weitere Vorgehensweise

Die Zielvariable Offenlegungsbereitschaft wurde über eine Offenlegung\_Skala mit vier Items operationalisiert, die die Akzeptanz konkreter Nachweismethoden abfragt:

- Abgabe von Screenshots oder Logs der KI-Interaktionen
- Klare Kennzeichnung aller KI-generierten Inhalte im Text (z. B. durch Markierung)
- Dokumentation der erstellten Prompts
- Nutzung von Plagiats- oder KI-Erkennungstools

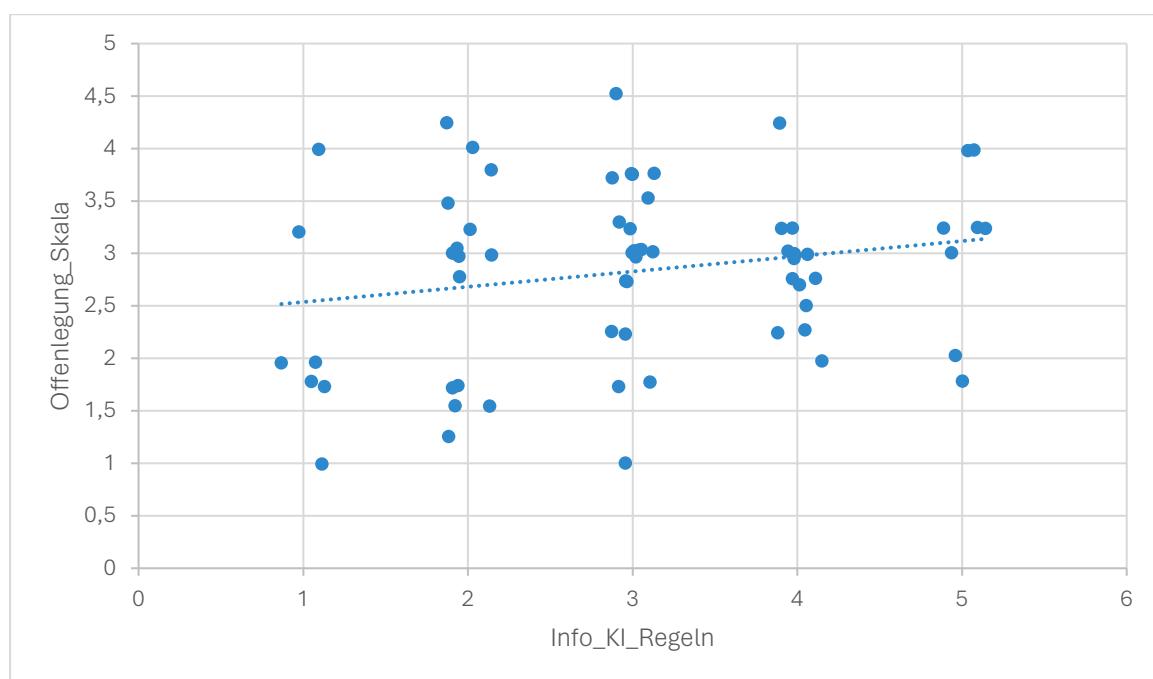
Diese Items leiten sich aus der Frage ab, inwieweit Studierende diese Maßnahmen als geeignet ansehen, um ihre Eigenleistung im KI-Kontext glaubhaft zu machen. Die Skala misst somit nicht nur eine generelle Offenlegungsabsicht, sondern die Bereitschaft zur konkreten Umsetzung formaler und inhaltlicher Nachweispflichten.

Die Reliabilität der Skala liegt bei Cronbach's Alpha = 0,649 und ist damit ausreichend. Weitere Variablen wurden nicht integriert, da frühere Analysen zeigten, dass die Erweiterung um zusätzliche Offenlegungsideen (z. B. Reflexionsbericht) die interne Konsistenz verringert hätte. Ziel war es, eine in sich konsistente Skala mit nachvollziehbarer inhaltlicher Kohärenz zu behalten, die explizit auf gängige Nachweisformen fokussiert.

Die anschließende lineare Regressionsanalyse nutzte die Variablen *Info\_KI\_Regeln* und *Offenlegung\_wenn\_andere* als Prädiktoren, um deren Einfluss auf die Offenlegungsbereitschaft der Studierenden zu untersuchen.

#### 4.4.3.2 Regressionsergebnisse

Die Regression zeigte insgesamt nur geringe Erklärungskraft ( $R^2 = 0,051$ ). Zwar zeigte sich ein tendenziell positiver Zusammenhang zwischen Regelwissen und Offenlegung ( $\beta = 0,206, p = 0,098$ ), dieser war jedoch nicht signifikant. Der Einfluss des sozialen Rückhalts durch Mitstudierende oder Lehrende (Offenlegung\_wenn\_andere) war noch geringer ( $\beta = 0,082, p = 0,505$ ).



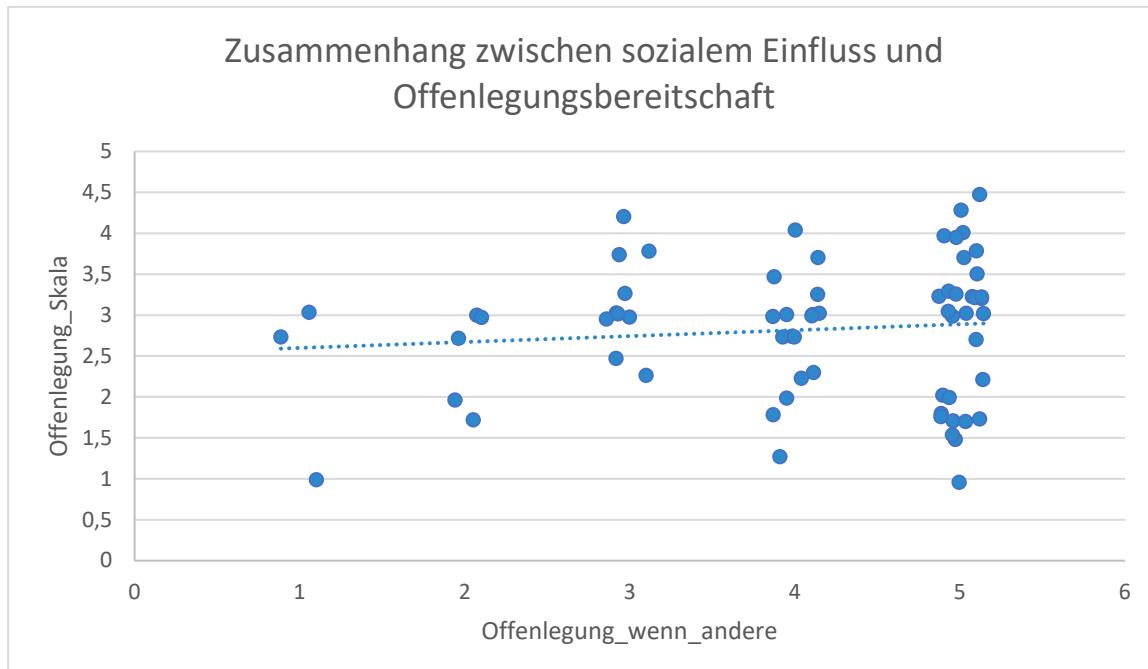


Abbildung 4.3: Zusammenhang zwischen sozialem Einfluss und Offenlegungsbereitschaft

Abbildung 3 zeigt, dass kein signifikanter Zusammenhang zwischen der Zustimmung zur Aussage „Ich würde KI-Nutzung offenlegen, wenn es andere auch tun“ und der tatsächlichen Bereitschaft zur Offenlegung besteht. Die Trendlinie verläuft nahezu horizontal, was durch die Regressionsanalyse bestätigt wird ( $\beta = 0,082$ ;  $p = 0,505$ ). Daraus lässt sich schließen, dass sozialer Rückhalt durch Mitstudierende oder Lehrende in diesem Fall kein relevanter Einflussfaktor für Offenlegung ist.

#### 4.4.4 Weitere Ergebnisse

Zur Ergänzung der Hypothesenprüfung wurden weitere individuelle Einflussfaktoren betrachtet, darunter Alter, Geschlecht, Semester, Hausarbeitenanzahl und ChatGPT-Erfahrung. Diese Analysen sollten klären, ob bestimmte Gruppen besonders zur Offenlegung oder zu einem strikten Eigenleistungsverständnis tendieren, unabhängig von Regelwissen oder Selbstwirksamkeit.

##### 4.4.4.1 Einfluss der soziodemografischen und themenspezifischen Variablen

Die Korrelationsanalyse zeigte keine signifikanten Beziehungen zwischen Alter, Studiensemester oder Zahl der Hausarbeiten pro Jahr und den Skalen „Eigenleistung“ bzw. „Offenlegung“ (alle  $p > 0,05$ ). Einzig das Alter korrelierte erwartungsgemäß positiv mit der Semesterzahl ( $r = 0,403$ ;  $p = 0,001$ ), was die Plausibilität der Daten bestätigt. Insgesamt sprechen die Ergebnisse für eine stimmige Messung, obwohl weitere Zusammenhänge ausblieben.

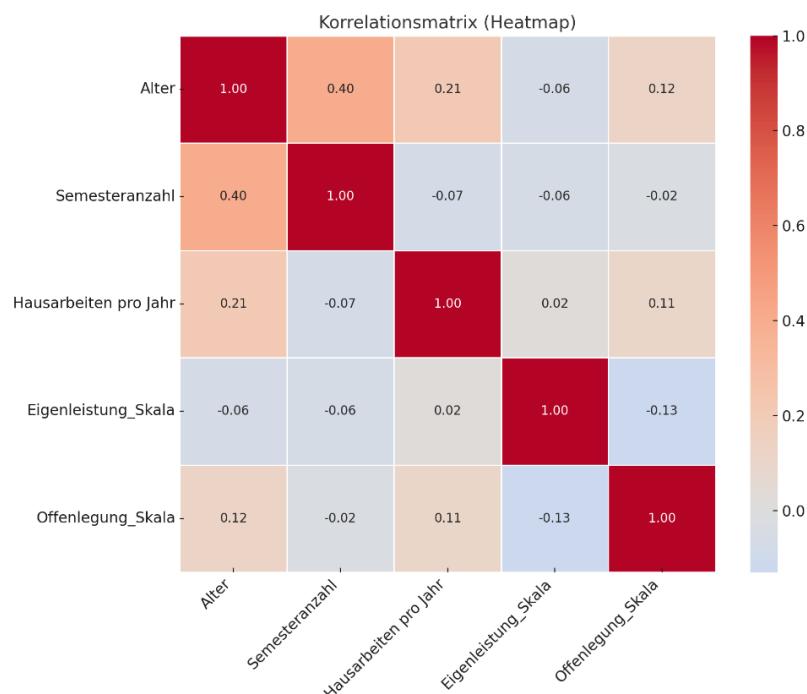


Abbildung 4.4: Zusammenhang zwischen Selbstwirksamkeit und Eigenleistungsverständnis

#### 4.4.4.2 Unterschiede nach Geschlecht – t-Tests für unabhängige Stichproben

Zur Analyse geschlechtsspezifischer Unterschiede bei Offenlegung und Eigenleistungsverständnis wurden t-Tests für unabhängige Stichproben durchgeführt. Beim Offenlegungsverhalten zeigte sich kein signifikanter Unterschied zwischen männlichen und weiblichen Studierenden ( $t(55,04) = -0,980$ ,  $p = 0,331$ ). Aufgrund signifikanter Varianzungleichheit (Levene-Test:  $p = 0,006$ ) kam der Welch-Test zum Einsatz. Frauen wiesen im Mittel eine leicht höhere Offenlegungsneigung auf, doch der Unterschied war statistisch nicht bedeutsam.

Beim Verständnis von Eigenleistung zeigte sich ein grenzwertiger, nicht signifikanter Unterschied zugunsten männlicher Befragter ( $t(63) = 1,949$ ,  $p = 0,056$ ). Männer erzielten im Mittel um 0,31 Punkte höhere Werte; das Konfidenzintervall [-0,008; 0,632] schloss jedoch die Null ein, sodass kein eindeutiger Unterschied festgestellt werden kann.

Die Angabe „divers“ ( $n = 1$ ) wurde aus methodischen Gründen ausgeschlossen, da der t-Test nur den Vergleich zweier Gruppen erlaubt und der Anteil zu gering für verlässliche Aussagen war.

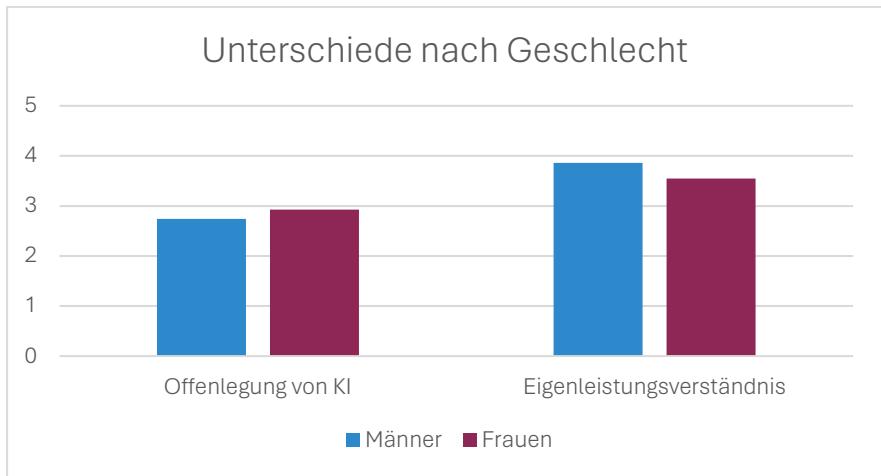


Abbildung 4.5: Unterschiede nach Geschlecht in Bezug auf Offenlegung und Eigenleistung

Das Balkendiagramm zeigt Mittelwerte männlicher und weiblicher Studierender zu Offenlegung von KI-Nutzung und Eigenleistungsverständnis. Frauen berichteten eine etwas höhere Offenlegungsbereitschaft, Männer ein leicht höheres Eigenleistungsverständnis. Die Unterschiede waren jedoch nicht signifikant ( $p > 0,05$ ).

#### 4.4.4.3 Einfluss von Erfahrung mit ChatGPT auf Eigenleistung und Offenlegung

Neben den Geschlechtsunterschieden wurde per linearer Regression untersucht, ob die selbst eingeschätzte ChatGPT-Erfahrung mit dem Eigenleistungsverständnis und der Offenlegungsbereitschaft zusammenhängt. Für die **Eigenleistung\_Skala** zeigte sich ein signifikanter positiver Zusammenhang ( $R = ,282$ ;  $R^2 = ,079$ ). Etwa 7,9 % der Varianz wurden durch den Prädiktor erklärt (korrigiertes  $R^2 = ,065$ ). Sowohl die Regressionskoeffizienten ( $B = ,137$ ;  $p = ,022$ ) als auch die Gesamt-ANOVA ( $F(1,64) = 5,524$ ;  $p = ,022$ ) waren signifikant.

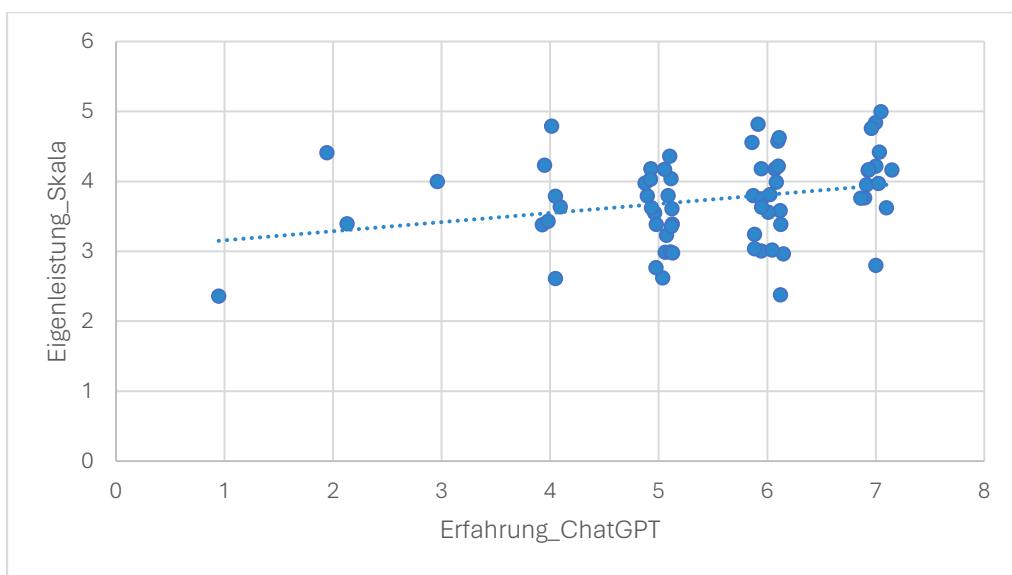


Abbildung 4.6: Zusammenhang zwischen der Erfahrung mit Nutzung von ChatGPT und des Eigenleistungsverständnisses

Für die *Offenlegung\_Skala* ergab die Analyse hingegen keinen signifikanten Zusammenhang zwischen ChatGPT-Erfahrung und der Bereitschaft, den KI-Einsatz offenzulegen ( $R = ,159$ ;  $R^2 = ,025$ ; korrigiertes  $R^2 = ,010$ ). Die erklärte Varianz ist gering, der Regressionskoeffizient ( $B = -,102$ ;  $p = ,204$ ) und das Gesamtmodell ( $F(1,64) = 1,650$ ;  $p = ,204$ ) waren nicht signifikant.

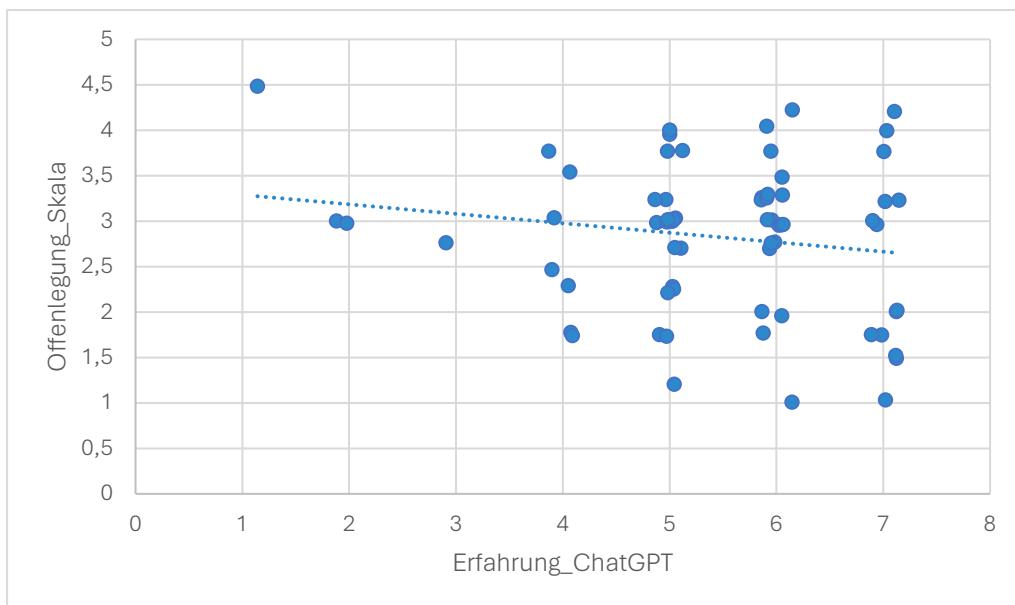


Abbildung 4.7: Zusammenhang zwischen der Erfahrung mit Nutzung von ChatGPT und der Offenlegungsbereitschaft

#### 4.4.5 Geeignete Nachweismethoden

Die Befragung zu geeigneten Nachweismaßnahmen für Eigenleistungen zeigt ein klares Meinungsbild: Am meisten Zustimmung erhielt der Reflexionsbericht ( $MW = 3,42$ ;  $SD = 0,93$ ), was auf eine als fair empfundene Methode hinweist.

Es folgen die Dokumentation von Prompts ( $MW = 3,11$ ) sowie die Kennzeichnung KI-generierter Inhalte im Text ( $MW = 2,97$ ), letzteres mit leicht erhöhter Streuung ( $SD = 1,22$ ), was auf geteilte Ansichten schließen lässt.

Weniger Zustimmung fanden technisch aufwändigere Maßnahmen wie spezielle Tools ( $MW = 2,85$ ) oder Screenshots ( $MW = 2,38$ ), die offenbar als umständlich oder weniger praktikabel wahrgenommen werden.

Item	Mittelwert (MW)	Standardabweichung (SD)
Nachweis_Reflexion	3,42	0,93
Nachweis_Prompts	3,11	1,11
Nachweis_Kennzeichnung	2,97	1,22
Nachweis_Tools	2,85	1,24
Nachweis_Screen	2,38	1,17

Tabelle 4.5: Nachweismethoden

## 4.5 Eigenleistung im Spiegel der Reflexion

In diesem Kapitel werden die zentralen Befunde der empirischen Untersuchung im Hinblick auf die aufgestellten Hypothesen und den theoretischen Rahmen interpretiert. Dabei werden die Ergebnisse kritisch eingeordnet, mit bestehenden Forschungsarbeiten in Beziehung gesetzt und hinsichtlich ihrer Aussagekraft und methodischen Begrenzungen reflektiert. Ziel ist es, die Befunde nicht nur zu interpretieren, sondern auch ihre Relevanz für wissenschaftliche Praxis und hochschuldidaktische Gestaltung aufzuzeigen.

### 4.5.1 Kritische Reflexion von Eigenleistung im Umgang mit KI

Die Ergebnisse stützen die Annahme, dass Studierende Eigenleistung nicht mehr primär durch die Vermeidung von KI-Nutzung definieren, sondern durch einen reflektierten und transparenten Umgang mit digitalen Tools. Die signifikante Korrelation zwischen der Nutzung von ChatGPT und der kritischen Reflexion von KI-generierten Inhalten weist darauf hin, dass Studierende aktiver zwischen maschineller Unterstützung und eigener Leistung differenzieren. Dieses Verständnis lässt sich im Licht des erweiterten Eigenleistungsbegriffs interpretieren, der im Theorieteil (Kap. 2.1) entwickelt wurde. Eigenleistung erfordert nicht nur Autorschaft im klassischen Sinne, sondern auch die Fähigkeit, den Beitrag generativer KI kritisch zu prüfen, zu transformieren und mit eigener Leistung abzugleichen (Ulrich & Brieden, 2021; Cromme, 2021, S. 65f.). In Anlehnung an Fleischmann, verlangt akademische Integrität im KI-Zeitalter eine reflektierte und regelgeleitete Nutzung, als Bestandteil wissenschaftlicher Eigenleistung (Fleischmann, 2023, S.23).

Besonders im akademischen Alltag bedeutet dies, Vermeidung wird nicht automatisch als Zeichen wissenschaftlicher Integrität gewertet, sondern könnte auch Ausdruck eines unreflektierten Umgangs mit neuen Technologien sein. Die Fähigkeit, die Grenze zwischen fremder maschineller Produktion und eigenem Denken zu markieren, stellt eine neue Form wissenschaftlicher Selbstverantwortung dar.

### 4.5.2 Selbstwirksamkeit und Haltung zu Eigenleistung

Die Regressionsergebnisse zur zweiten Hypothese zeigen, dass ein höheres Maß an Selbstwirksamkeit mit einem strengeren Verständnis von Eigenleistung einhergeht, dieser Zusammenhang jedoch nicht mehr besteht, wenn berücksichtigt wird, wie positiv Studierende KI-Ergebnisse als eigene Leistung ansehen. Dieser Befund lässt sich im Rückgriff auf Bandura (1977) so deuten: Selbstwirksamkeit beeinflusst nicht nur das Verhalten, sondern auch die zugrunde

liegenden normativen Maßstäbe (Bandura, 1977, S. 193). Dass der Effekt der Selbstwirksamkeit bei Kontrolle der KI-Haltung verschwindet, deutet darauf hin, dass ethische Überzeugungen den Zusammenhang beeinflussen. Dies stimmt mit den Ergebnissen von Sallam et al. (2024, S. 8-10) überein, die zeigen, dass Selbstwirksamkeit im Sinne des Vertrauens in die eigene Kompetenz einen wesentlichen Einfluss auf die Nutzung und Bewertung von KI hat. Unsere Ergebnisse erweitern diese Perspektive, indem sie zeigen, dass ein hohes Maß an Selbstwirksamkeit häufig mit einem strengeren Verständnis von Eigenleistung verbunden ist, allerdings nur dann, wenn die Studierenden nicht gleichzeitig eine positive Haltung gegenüber KI vertreten. Studierende mit hoher Selbstwirksamkeit, die KI als legitimen Bestandteil akademischer Leistung anerkennen, wenden ihre kognitiven Ressourcen möglicherweise nicht zur Abgrenzung, sondern zur Integration von KI an. Damit lässt sich schlussfolgern, dass Eigenleistungsverständnis nicht allein durch psychologische Variablen wie z. B. Kompetenzgefühl geprägt wird, sondern durch ein komplexes Zusammenspiel individueller Haltungen, institutioneller Normen und technologischer Akzeptanz.

#### **4.5.3 Institutionelle Regelkenntnis und sozialer Einfluss**

Hypothese 3 wurde nur tendenziell gestützt. Zwar zeigt sich ein schwacher positiver Zusammenhang zwischen Regelwissen und Offenlegungsbereitschaft, dieser ist jedoch nicht signifikant. Noch schwächer war der Einfluss des sozialen Rückhalts durch Lehrende oder Mitstudierende. Diese Ergebnisse deuten darauf hin, dass individuelle strategische Überlegungen, etwa zur Wahrung von Vorteilen oder zur Vermeidung von Konflikten, eine größere Rolle spielen als soziale oder institutionelle Orientierung. Studierende scheinen sich trotz Wissen um Regelungen zurückhaltend zu verhalten, möglicherweise aus taktischen Gründen, zum Beispiel um eine „unsichtbare“ Nutzung von KI nicht offenlegen zu müssen oder um kritische Rückfragen zu vermeiden. Der theoretische Bezug zum UTAUT-Modell zeigt, dass „social influence“ allein nicht genügt (Venkatesh et al., 2003, S. 451), wenn institutionelle Normen nicht klar kommuniziert oder durchgesetzt werden. Wenn Regeln unklar bleiben, bietet auch das Verhalten anderer kaum Orientierung und kann etablierte Standards infrage stellen. Die daraus entstehende Lücke zwischen Regelkenntnis und Umsetzung verweist auf informelle Gewohnheiten und unausgesprochene Praktiken, die durch Unsicherheit, Routinen oder mangelnde Sanktionierung geprägt sind.

#### **4.5.4 Präferenzen bei Nachweismethoden und hochschuldidaktische Implikationen**

Die klaren Befunde zur Frage geeigneter Nachweismethoden zeigen, dass Studierende vor allem reflexionsbasierte und erklärende Formate bevorzugen, insbesondere den Reflexionsbericht. Dies passt zu einem erweiterten Leistungsverständnis, das Eigenleistung nicht technisch misst, sondern durch Argumentation, Einordnung und persönliches Verständnis sichtbar macht (Fleischmann, 2025, S. 237). Wenig Zustimmung finden technische Nachweise wie Screenshots oder spezielle Tools. Diese werden offenbar als wenig praktikabel oder als Vertrauensbruch empfunden. Die Hochschuldidaktik sollte daraus ableiten, dass dokumentierte Eigenleistung nicht über Kontrolle, sondern über didaktisch sinnvolle Formate gestaltet werden sollte, etwa durch Integration von Reflexionsfragen, Prompt-Dokumentationen oder mündliche Erläuterungen (Hof, 2025, S.633).

Die Ergebnisse passen zu Beobachtungen aus der Studie von Garrel & Mayer (2025, S. 13), die zeigen, dass viele Hochschulen auf reflexive Formate setzen, um Eigenleistung im KI-Kontext zu dokumentieren. Die hohe Zustimmung zu Reflexionsberichten in unserer Studie legt nahe, dass diese Formate von Studierenden als besonders glaubwürdig und lernförderlich empfunden werden.

#### **4.5.5 Methodenkritik und Kontextualisierung**

Die Ergebnisse sind vor dem Hintergrund methodischer Einschränkungen zu interpretieren. Die Stichprobe umfasst lediglich 66 Studierende eines Studiengangs an einer Hochschule. Die nicht-probabilistische Gelegenheitsstichprobe begrenzt die Repräsentativität, auch wenn die Rücklaufquote von 13,2 % respektabel ist. Außerdem liegt ein rein quantitatives Design vor, das keine tiefergehende Einsicht in Motive, Konflikte oder Bedeutungszuschreibungen erlaubt. Weitere Forschung sollte qualitative Verfahren einbeziehen, um subjektive Deutungen von Eigenleistung, Regelverständnis und KI-Verwendung detaillierter zu erfassen. Auch ein Vergleich mit anderen Studiengängen oder Hochschulformen könnte Aufschluss darüber geben, ob sich die beschriebenen Muster verallgemeinern lassen oder stark kontextabhängig sind (Garrel & Mayer, 2025; Gottschling, Seidl & Vonhof, 2024).

### **4.6 Neuverhandlung wissenschaftlicher Eigenleistung im KI-Zeitalter**

Die vorliegende Untersuchung beschäftigte sich mit der Frage, wie Studierende Eigenleistung im Zeitalter generativer KI definieren und wie sie sicherstellen, diese bei wissenschaftlichen Arbeiten glaubwürdig nachweisen zu können. Ausgehend von einem erweiterten Eigenleistungsbegriff, der neben Selbstständigkeit auch Transparenz und Reflexion im Umgang mit KI umfasst, wurden drei Hypothesen formuliert und auf Basis einer quantitativen Online-Befragung unter Studierenden der Wirtschaftsinformatik an der Frankfurt University of Applied Sciences geprüft.

Die Ergebnisse zeigen, dass Eigenleistung nicht mehr primär durch den Verzicht auf digitale Tools definiert wird. Stattdessen verstehen viele Studierende den reflektierten, kritischen und transparenten Einsatz von KI als Teil ihrer Eigenleistung. Dies stützt Hypothese 1, zwischen maschinellem Textgenerierung und eigenem wissenschaftlichen Beitrag wird bewusst unterschieden, vor allem von jenen, die KI regelmäßig und mit einer gewissen Vertrautheit nutzen. Damit etabliert sich ein neues Leistungsverständnis, das nicht die bloße Ablehnung, sondern die Fähigkeit zur kritischen Integration und Weiterverarbeitung von KI-Inhalten betont.

Bezogen auf Hypothese 2 zeigt sich ein differenziertes Bild. Studierende mit hoher Selbstwirksamkeit neigen zu einem strengeren Verständnis von Eigenleistung. Dieser Zusammenhang verliert jedoch an statistischer Bedeutung, wenn zugleich eine positive Haltung zur KI-Nutzung als legitime Eigenleistung besteht. Eigenleistungsverständnis ist somit nicht allein psychologisch durch Kompetenzgefühl bedingt, sondern durch das Zusammenspiel individueller Werthaltungen und technischer Akzeptanz geprägt.

Hypothese 3 wurde nur schwach gestützt. Zwar zeigen die Daten einen leichten positiven Zusammenhang zwischen Regelkenntnis und Offenlegung, dieser ist jedoch nicht signifikant. Noch geringer ist der Einfluss sozialer Orientierung an Lehrenden oder Mitstudierenden. Daraus lässt sich ableiten, dass strategische Überlegungen und individuelle Einschätzungen eine größere Rolle spielen als formale oder soziale Vorgaben. In einem Umfeld mit unklaren oder uneinheitlich kommunizierten Regeln orientieren sich Studierende offenbar stärker an pragmatischen Lösungen.

Übergreifend lässt sich festhalten, Eigenleistung wird durch reflektierte Auseinandersetzung definiert, nicht durch Vermeidung. Studierende erkennen, dass die Glaubwürdigkeit wissenschaftlicher Leistung im KI-Kontext vor allem durch Transparenz, kritische Bewertung und nachvollziehbare Dokumentation entsteht. Gleichzeitig zeigen die Ergebnisse, dass individuelle Einstellungen das Offenlegungsverhalten stärker prägen als institutionelle Regelwerke.

Reflexionsorientierte Nachweismethoden, etwa in Form von Berichten oder der Dokumentation von Prompts, stoßen unter Studierenden auf hohe Akzeptanz. Technisch-kontrollierende Maßnahmen wie Screenshots oder automatisierte Tools werden hingegen skeptisch bewertet. Hochschulen sind daher gefordert, Nachweismethoden zu entwickeln, die lernförderlich und gleichzeitig glaubwürdig sind, etwa durch die didaktische Einbettung reflexiver Formate statt durch Kontrolle.

Gleichzeitig wird deutlich, wo weiterer Forschungsbedarf besteht. Die vorliegende Untersuchung basiert auf einem rein quantitativen Design und einer begrenzten Stichprobe. Künftige Studien sollten qualitative Methoden wie Interviews oder Fokusgruppen nutzen, um subjektive Perspektiven auf Eigenleistung, Offenlegung und Regelverständnis tiefergehend zu erfassen. Besonders in Graubereichen, in denen Normen unklar sind, können solche Zugänge ein differenzierteres Bild liefern.

Auch disziplinäre und institutionelle Vergleichsstudien sind lohnenswert. Die Ergebnisse dieser Arbeit beziehen sich auf einen wirtschaftsinformatischen Studiengang an einer Hochschule für angewandte Wissenschaften. Unterschiede zu anderen Fachkulturen oder Hochschultypen, etwa Universitäten oder künstlerisch geprägten Disziplinen, könnten neue Einsichten zur normativen Deutung von Eigenleistung im KI-Kontext liefern. Ebenso ist ein internationaler Vergleich sinnvoll, da Regelungen, Erwartungen und technologische Infrastruktur weltweit stark variieren.

Abschließend stellt sich die grundlegende Frage, ob wissenschaftliche Eigenleistung im Zeitalter generativer KI neu definiert werden muss: nicht mehr nur als Leistung trotz KI, sondern als Leistung durch den reflektierten Umgang mit ihr. Es geht dabei weniger um die strikte Trennung von menschlicher und maschineller Beteiligung, sondern vielmehr um deren bewusste Verbindung, um die Fähigkeit, maschinelle Unterstützung kritisch zu steuern, produktiv einzuordnen und verantwortungsvoll in eigene Denkprozesse zu integrieren.

Diese Transformation gelingt nur, wenn Forschung, Lehre und Prüfung gemeinsam gedacht werden, als Teile eines wissenschaftlichen Bildungsprozesses, der Eigenverantwortung, Transparenz und technologische Mündigkeit ins Zentrum stellt. In diesem Sinne gewinnt auch ein klassischer Bildungsbegriff erneut an Relevanz: Bildung als „Selbstbestimmungs-, Mitbestimmungs- und Solidaritätsfähigkeit“ (Klafki, 1985, S. 17), gerade im reflektierten Umgang mit neuen Technologien.

## 4.7 Literaturverzeichnis

- Almassaad, A., Alajlan, H., & Alebaikan, R. (2024). Student perceptions of generative artificial intelligence: Investigating utilization, benefits, and challenges in higher education. *Systems*, 12(10), Article 385. <https://doi.org/10.3390/systems12100385>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Cromme, R. (2021). Fordern, fördern, füttern – studentische Eigenständigkeit aktivieren. In J. Noller et al. (Hrsg.), *Studierendenzentrierte Hochschullehre: Perspektiven der Hochschuldidaktik* (S. 65–80). Springer. [https://doi.org/10.1007/978-3-658-32205-2\\_4](https://doi.org/10.1007/978-3-658-32205-2_4)
- Duden.de. (2024). Eigenleistung. <https://www.duden.de/node/37294/revision/1339717>
- Eigenständigkeitserklärung. (2023). Eigenständigkeitserklärung. [https://www.frankfurt-university.de/fileadmin/standard/Studium/Pruefungsanmeldung/Eigenstaendigkeitserklaerung\\_barrierefrei.pdf](https://www.frankfurt-university.de/fileadmin/standard/Studium/Pruefungsanmeldung/Eigenstaendigkeitserklaerung_barrierefrei.pdf)
- Faraon, M., Rönkkö, K., Milrad, M., & Tsui, E. (2025). International perspectives on artificial intelligence in higher education: An explorative study of students' intention to use ChatGPT across the Nordic countries and the USA. *Education and Information Technologies*, 30(4), 1234–1250. <https://doi.org/10.1007/s10639-025-13492-x>

- Fleischmann, A. (2025). Über den Einfluss von künstlicher Intelligenz in der Hochschullehre. In H. Lackner, J. Cai & Q. Wang (Hrsg.), *Jahrbuch Angewandte Hochschulbildung 2023: Deutsch-Chinesische Perspektiven und Diskurse* (S. 226–238). Springer VS.
- Fleischmann, K. (2023). ChatGPT und die Zukunft der Hochschullehre: Chancen, Risiken und Perspektiven. In A. Schulz, K. Reuter & M. Weis (Hrsg.), *Künstliche Intelligenz in der Bildung: Diskurse und Perspektiven* (S. 17–26). Springer VS.
- Frankfurt University of Applied Sciences. (2023a). Allgemeine Bestimmungen für Prüfungsordnungen mit den Abschlüssen Bachelor und Master (AB Bachelor/Master) [Leseversion vom 21. Juni 2023]. <https://www.frankfurt-university.de/de/aktuelles/amtliche-mitteilungen/studien-pruefungs-und-entgeltordnungen/>
- Garrel, J. von, & Mayer, J. (2025). Künstliche Intelligenz im Studium: Eine quantitative Längsschnittstudie zur Nutzung KI-basierter Tools durch Studierende (2023 & 2025). Hochschulforum Digitalisierung. [https://doi.org/10.48444/h\\_docs-pub-533](https://doi.org/10.48444/h_docs-pub-533)
- Gesetze im Internet. (2023). Urheberrechtsgesetz (UrhG), § 2 Begriffsbestimmung. Bundesministerium der Justiz. [https://www.gesetze-im-internet.de/urhg/\\_2.html](https://www.gesetze-im-internet.de/urhg/_2.html)
- Gottschling, M., Seidl, F., & Vonhof, S. (2024). Nutzung von KI-Tools durch Studierende: Eine exemplarische Untersuchung studentischer Nutzungsszenarien. *die hochschullehre*, 10, 123–136. <https://doi.org/10.3278/HSL2411W>
- Gumprecht, S., Meyer, S., Peppel, L., Flick, M., Pukal, K., & Kuhn, C. (2024). Lehren, Lernen und Schreiben mit generativen KI-Anwendungen: Empfehlungen und Impulse (Version 4.2). Frankfurt University of Applied Sciences. [https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren\\_Lernen\\_und\\_Schreiben\\_mit\\_generativen\\_KIAngewandten\\_BeSt\\_BiB\\_Schreibwerkstatt\\_V4.2.pdf](https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren_Lernen_und_Schreiben_mit_generativen_KIAngewandten_BeSt_BiB_Schreibwerkstatt_V4.2.pdf)
- Hessisches Hochschulgesetz (HessHG). (2021). Gesetz über die Hochschulen des Landes Hessen. GVBl. 2021, S. 931. Zuletzt geändert durch Gesetz vom 18. Juli 2023 (GVBl. S. 566). <https://www.rv.hessenrecht.hessen.de/bshe/document/jlr-HSchulGHE2021V1>
- Hof, H.-J. (2025). Digitalisierung in Hochschulen. In L. Fend & J. Hofmann (Hrsg.), *Digitalisierung in Industrie-, Handels- und Dienstleistungsunternehmen: Konzepte – Lösungen – Beispiele* (4. Aufl., S. 616–634). Springer Gabler.
- Klafki, W. (1985). Neue Studien zur Bildungstheorie und Didaktik. Weinheim: Beltz.
- Mittelstraß, J. (1990). Von der Freiheit der Forschung und der Verantwortung des Wissenschaftlers. *The Science of Nature*, 77(4), 149–157. <https://doi.org/10.1007/BF01131156>
- Saldan, P., & Leschke, J. (2024). Learning Analytics und Künstliche Intelligenz in Studium und Lehre: Erfahrungen und Schlussfolgerungen aus einer hochschulweiten Erprobung. Springer-Verlag.
- Sallam, M., Alhadidi, M., Al-Mahzoum, K., & Al-Salahat, K. (2024). ChatGPT usage and attitudes are driven by perceptions of usefulness, ease of use, risks, and psycho-social impact: A study among university students in the UAE. *Frontiers in Education*, 9, Article 1414758. <https://doi.org/10.3389/feduc.2024.1414758>
- Ulrich, I., & Brieden, M. (2021). Studierendenzentrierte Hochschullehre aus lernpsychologischer Sicht. In J. Noller et al. (Hrsg.), *Studierendenzentrierte Hochschullehre: Perspektiven der Hochschuldidaktik* (S. 3–22). Springer. [https://doi.org/10.1007/978-3-658-32205-2\\_1](https://doi.org/10.1007/978-3-658-32205-2_1)
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1–27. <https://doi.org/10.1186/s41239-019-0171-0>

## 5 KI in schriftlichen Prüfungsformaten

### *KI und ihre Implikationen für schriftliche Prüfungsformate an Hochschulen*

Madinah Mehrabuddin

Hamdan Razaq

Alia Anjum Rizvi

#### **Abstract**

Generative KI-Tools wie ChatGPT verändern die Bedingungen für Hausarbeiten an Hochschulen. Diese Studie untersucht, wie Studierende deren Einsatz wahrnehmen und auf neue Anforderungen sowie hochschulische Maßnahmen reagieren.

Dazu wurde eine quantitative Online-Befragung unter 66 Studierenden der Frankfurt University of Applied Sciences durchgeführt. Die Auswertung erfolgte mithilfe deskriptiver Statistik, Regressionsanalysen, nicht-parametrischer Tests und Clusteranalysen.

Die Ergebnisse zeigen, dass 85 % der Studierenden KI-Tools als Unterstützung im Schreibprozess sehen. Inhaltliche und methodische Anforderungen werden wichtiger eingeschätzt als formale Aspekte wie Struktur oder Sprache. Konzepte zur kontrollierten Integration von KI stoßen dabei auf deutlich mehr Zustimmung als restriktive Maßnahmen. Zudem wird erkennbar, dass Studierende mit mehr Erfahrung und positiver Haltung KI stärker als Mittel zur Qualitätssteigerung und Prozessoptimierung bewerten, während skeptische Personen eher negative Effekte auf Reflexion und Eigenständigkeit betonen.

Die Befunde liefern erste empirische Hinweise darauf, dass Prüfungsformate angepasst werden sollten, um Eigenleistung trotz KI-Nutzung nachvollziehbar zu machen. Besonders wichtig sind dabei methodisches Vorgehen, kritische Auseinandersetzung und Reflexion. Studierende lehnen Verbote ab, befürworten aber eine bedingte Zulassung von KI sowie ergänzende Formate wie Projekte oder mündliche Prüfungen, um individuelle Leistungen sichtbar zu halten.

**Schlagwörter:** generative KI, Hochschulbildung, Hausarbeiten, Prüfungsformate, akademische Integrität, studentische Perspektive, Hochschulmaßnahmen

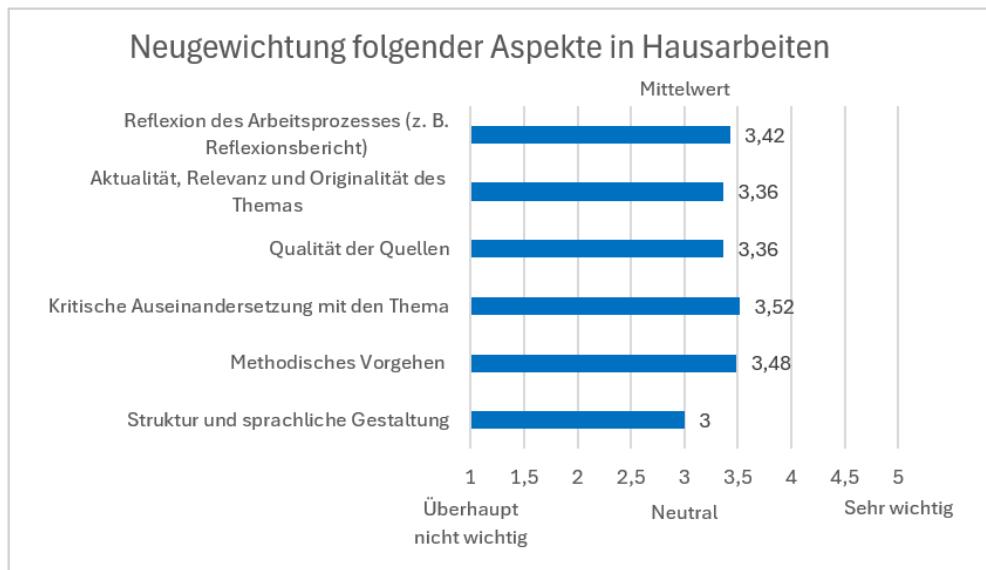


Abbildung 5.1: Neugewichtung von Aspekten in Hausarbeiten

## 5.1 Prüfungskultur im Wandel: Zur Notwendigkeit dieser Untersuchung

### 5.1.1 Relevanz des Themas

Die zunehmende Verfügbarkeit generativer KI-Anwendungen wie ChatGPT verändert die Bedingungen des wissenschaftlichen Arbeitens an Hochschulen. Laut einer aktuellen Studie des Digitalverbands Bitkom hat bereits rund die Hälfte der deutschen Studierenden entsprechende Tools im Studium eingesetzt, vor allem zur Unterstützung bei der Formulierung, Strukturierung und sprachlichen Überarbeitung wissenschaftlicher Texte (Bitkom, 2024).

Diese Entwicklung stellt etablierte Prüfungsformate, insbesondere die schriftliche Hausarbeit, vor neue Herausforderungen. Hausarbeiten beruhen traditionell auf der Annahme individueller Leistungserbringung und darauf, dass Studierende ihre Texte eigenständig und ohne inhaltlich unterstützende Systeme verfassen. Der wachsende Einsatz generativer KI relativiert diese Grundlage und wirft grundlegende Fragen nach Fairness, akademischer Integrität und pädagogischer Zielsetzung auf.

Der institutionelle Umgang mit dieser Thematik ist bislang uneinheitlich. Während einige Hochschulen restriktive Regelungen formulieren, entwickeln andere erste Leitlinien zur gezielten Integration von KI in Prüfungsformate.

Vor diesem Hintergrund erscheint eine systematische Untersuchung der Wahrnehmungen von Studierenden zur Rolle generativer KI in schriftlichen Leistungsnachweisen sinnvoll.

### 5.1.2 Zielsetzung und Forschungsfragen

Ziel dieser Arbeit ist es, studentische Einschätzungen zur Rolle generativer KI im Kontext schriftlicher Prüfungsformate zu erfassen. Im Zentrum stehen dabei zwei Fragen, die im Rahmen einer quantitativen Online-Befragung beantwortet wurden:

- 1 Welche Auswirkungen hat die Nutzung von KI auf die Anforderungen klassischer Hausarbeiten aus der Sicht von Studierenden?
- 2 Wie bewerten Studierende die Maßnahmen deutscher Hochschulen im Umgang mit KI in schriftlichen Prüfungsformaten?

### **5.1.3 Struktur der Untersuchung**

Kapitel 2 führt in zentrale Begriffe und theoretische Grundlagen ein und bietet einen Überblick über den bisherigen Forschungsstand zum Thema KI und schriftliche Prüfungsformate. Anschließend beschreibt Kapitel 3 das methodische Vorgehen, einschließlich der Gestaltung der Umfrage, der Datenerhebung sowie der eingesetzten statistischen Auswertungsverfahren. Die Ergebnisse der Befragung werden in Kapitel 4 dargestellt. Darauf aufbauend ordnet Kapitel 5 diese Ergebnisse in den theoretischen und empirischen Kontext ein und diskutiert zentrale Befunde sowie mögliche Einschränkungen der Studie. Abschließend fasst Kapitel 6 die wichtigsten Erkenntnisse zusammen und gibt einen Ausblick auf weiterführende Fragestellungen.

## **5.2 KI in Hochschulen im Spiegel von Praxis und Forschung**

### **5.2.1 Institutionelle Regelungen und rechtlicher Rahmen**

Rechtlich wirft der Einsatz generativer KI im Hochschulkontext eine Reihe neuer Fragestellungen auf. Dazu zählen insbesondere die Fragen nach der Urheberschaft, der Definition von Täuschung und der Nachvollziehbarkeit erbrachter Prüfungsleistungen. Da bislang keine spezifischen gesetzlichen Regelungen auf Bundesebene vorliegen, wird der Umgang mit KI-gestützten Leistungen derzeit überwiegend durch hochschulinterne Richtlinien bestimmt. Die HIS-HE-Studie dokumentiert anhand von Praxisbeispielen, wie Hochschulen generative KI im Lehrkontext bereits einsetzen und dafür praktische Handlungsstrategien entwickeln (HIS-HE, 2025, S. 6).

Der institutionelle Umgang mit generativer KI variiert dabei erheblich. Während einige Hochschulen klare Einschränkungen oder Verbote formulieren, erarbeiten andere erste Leitlinien zum reflektierten Einsatz in Lehre und Prüfungen. Für die vergleichende Analyse institutioneller Reaktionen wurden gezielt Handlungsempfehlungen und Richtlinien von 16 ausgewählten Hochschulen einbezogen. Berücksichtigt wurden insbesondere größere, öffentlich sichtbare und forschungsstarke Universitäten, darunter einige Technischen Universitäten. Aufgrund ihrer technischen Ausrichtung und bildungspolitischen Bedeutung ist anzunehmen, dass gerade diese Einrichtungen frühzeitig und systematisch auf die Herausforderungen der KI-Nutzung in Prüfungsprozessen reagieren. Ergänzend wurden praxisorientierte Fachhochschulen einbezogen, die durch ihre hohe Studierendenzahl und enge Anbindung an den Berufsalltag zusätzliche Perspektiven auf den digitalen Wandel eröffnen.

Hochschulen	
Fernuniversität in Hagen	Technische Universität Braunschweig
FOM-Hochschule	Technische Universität Darmstadt
IU Internationale Hochschule	Technische Universität Dresden
Johann Wolfgang-Goethe-Universität	Technische Universität München
Leibniz Universität Hannover	Technische Universität Hamburg
Universität Münster	Technische Universität Berlin
Rheinisch-Westfälische Technische Hochschule Aachen	Ludwig-Maximilians-Universität München
Hochschule Fulda	Universität Köln

Tabelle 5.1: Übersicht der analysierten Hochschulen

Ein Großteil der analysierten Hochschulen erlaubt den Einsatz generativer KI grundsätzlich, verknüpft diesen jedoch mit Bedingungen wie der transparenten Kennzeichnung, dem Nachweis eigenständiger Leistung und der Einhaltung wissenschaftlicher Standards. Hochschulen wie die LMU München oder die FernUniversität in Hagen fordern darüber hinaus die Dokumentation verwendeter Prompts oder eine methodische Reflexion des KI-Einsatzes. In mehreren Fällen, wie an der Universität zu Köln, der RWTH Aachen oder der Technischen Universität München, liegt die Entscheidung über die Zulässigkeit beim jeweiligen Lehrpersonal und hängt vom Prüfungsziel sowie dem erwarteten Kompetenznachweis ab.

Demgegenüber äußern sich einige Hochschulen deutlich kritischer. So verweist die Goethe-Universität Frankfurt auf die begrenzte Verlässlichkeit generativer KI und betont, dass Hausarbeiten weiterhin als eigenständige Leistungen zu erbringen seien. Auch die Universität Münster schließt eine Nutzung dort aus, wo sprachliche Eigenleistung oder kreative Ausdrucksfähigkeit Prüfungsbestandteil sind. Ein Überblick von SciBbr (2023), der über 100 Hochschulen umfasst, bestätigt diese Vielfalt. Neben offenen und lehrenden basierten Regelungen finden sich auch vereinzelt generelle Verbote des KI-Einsatzes, etwa ohne explizite Genehmigung oder in prüfungsrelevanten Kontexten wie Abschlussarbeiten.

Insgesamt zeigt sich, dass der Einsatz generativer KI an Hochschulen nicht grundsätzlich verboten ist, jedoch zunehmend an Kriterien wie Transparenz, Zielorientierung und didaktische Einbettung geknüpft wird.

## 5.2.2 Aktueller Forschungsstand

Der Forschungsstand zur Nutzung generativer KI in der Hochschulbildung ist noch begrenzt, da dieses Forschungsfeld erst im Entstehen begriffen ist. Erste Studien, etwa von Chan und Hu sowie Johnston et al. deuten darauf hin, dass Studierende KI-Tools vor allem pragmatisch verwenden. Sie setzen diese primär zur Unterstützung ein, nicht jedoch als vollständigen Ersatz für eigene Leistungen. Zugleich bestehen Unsicherheiten über deren Auswirkungen auf wissenschaftliche Prüfungsformate. (Chan und Hu, 2023, S.8; Johnston et al, 2024, S.1)

Insbesondere im deutschsprachigen Raum fehlen systematische empirische Untersuchungen zur Rolle generativer KI in schriftlichen Prüfungen wie Hausarbeiten. Forschungslücken betreffen vor allem die Auswirkungen der KI-Nutzung auf Bewertungskriterien und inhaltliche Zielsetzungen. Auch

die studentische Perspektive auf didaktische Eignung und Zukunftsfähigkeit schriftlicher Leistungsnachweise ist bislang kaum untersucht. Gleiches gilt für ihre Erwartungen und Forderungen an den hochschulischen Umgang mit KI.

Die vorliegende Arbeit will zum aktuellen Forschungsstand beitragen, indem sie studentische Einschätzungen zu Prüfungsanforderungen und hochschulischen Maßnahmen im KI-Kontext systematisch erfasst und analysiert.

### **5.3 Von der Theorie zur Datengrundlage**

Die Auswertung der erhobenen Daten erfolgte mit Microsoft Excel und Google Colab. Zunächst wurden deskriptive Statistiken berechnet, um zentrale Merkmale der Stichprobe darzustellen.

Im Anschluss kamen verschiedene inferenzstatistische Verfahren zur Anwendung, die sich am jeweiligen Anwendungsfall der Variablen orientierten. Zunächst wurde eine Korrelationsanalyse durchgeführt. Merkmalsbereiche mit einem Korrelationskoeffizienten von mindestens  $r = 0,30$  wurden vertieft analysiert.

Da es sich überwiegend um ordinalskalierte Daten auf Basis von Likert-Skalen handelte und eine Normalverteilung nicht vorausgesetzt werden konnte, wurden vorrangig geeignete nicht-parametrische Verfahren verwendet. Zur Untersuchung von Gruppenunterschieden kam der Kruskal-Wallis-Test zur Anwendung, sofern mindestens drei Gruppen vorlagen. Bei dichotomen Merkmalen wie dem Geschlecht wurde ergänzend der Mann-Whitney-U-Test eingesetzt.

Zur Ermittlung potenzieller Einflussfaktoren auf bestimmte Bewertungen wurden ordinale Regressionsanalysen eingesetzt. Abschließend wurde eine explorative Clusteranalyse nach dem K-Means-Verfahren durchgeführt, um typische Muster in der studentischen Haltung und Nutzungspraxis generativer KI zu identifizieren. Für weiterführende methodische Details wird auf das übergeordnete Methodikkapitel verwiesen.

### **5.4 Empirische Einblicke in studentische Perspektiven**

Das Kapitel fasst die zentralen Befunde der quantitativen Befragung zusammen. Es umfasst deskriptive Analysen sowie weitere statistische Auswertungen als Basis für die Diskussion.

## 5.4.1 Deskriptive Ergebnisse der Befragung

### 5.4.1.1 Einfluss von KI

Die Tabelle zeigt, inwieweit Studierende Aussagen zu den Auswirkungen von ChatGPT auf klassische Hausarbeiten zustimmen. Der Einsatz generativer KI wird insgesamt als unterstützend im Schreibprozess bewertet.

Aussage	MW	SD
KI verbessert Qualität und Arbeitsprozess	4,26	0,97
KI erhöht Erwartungen	3,61	1,07
KI mindert die Aussagekraft	3,30	1,16
KI beeinträchtigt kognitive Leistung	3,21	1,27

Tabelle 5.2: Einschätzung der Auswirkungen von KI auf Hausarbeiten

Die höchste Zustimmung gilt der Aussage, dass KI die Qualität der Arbeit und den Arbeitsprozess verbessert. Zur Aussagekraft individueller Leistungen und zu möglichen kognitiven Beeinträchtigungen äußern sich die Befragten hingegen zurückhaltender.

### 5.4.1.2 Anforderungen an Hausarbeiten

Im Hinblick auf die Anforderungen an wissenschaftliche Hausarbeiten im Kontext generativer KI betonen die Befragten inhaltliche und analytische Aspekte.

Aussage	MW	SD
Kritische Auseinandersetzung	3,52	0,93
Methodik	3,48	0,75
Reflexionsbericht	3,42	0,93
KI-Kennzeichnung	3,39	1,07
Themenrelevanz	3,36	0,67
Quellenqualität	3,36	0,91
Dokumentation der Prompts	3,11	1,11
Struktur und Sprache	3,00	0,88
KI-Erkennungstools	2,85	1,24

Tabelle 5.3: Bewertung von Anforderungen im KI-Kontext

Die kritische Auseinandersetzung mit dem Thema sowie das methodische Vorgehen werden im Vergleich zu formalen Kriterien wie Struktur oder sprachlicher Gestaltung höher bewertet. Auch die

Reflexion über den KI-Einsatz wird als relevant eingeschätzt, nimmt jedoch eine nachgeordnete Rolle ein.

Die Dokumentation der Prompts und technische Kontrollmechanismen wie automatische Erkennungstools stoßen dagegen auf geringere Akzeptanz.

#### **5.4.1.3 Hochschulischer Regelungen**

Die Tabelle zeigt, wie die Teilnehmenden verschiedene Maßnahmen zur Regulierung von KI-Nutzung sowie alternative Prüfungsformate einschätzen.

Aussage	MW	SD
Bedingte Zulassung KI	3,50	1,09
Ergänzung durch Alternativen	3,09	1,01
Reduktion von Hausarbeiten	2,97	1,25
Ersetzung von Hausarbeiten	2,59	1,38
KI-Verbot	1,67	0,94

Tabelle 5.4: Zustimmung zu hochschulischen Maßnahmen

Die Aussage, dass KI bedingt zugelassen oder gefördert werden sollte, erhielt die höchste Zustimmung. Eine generelle Ablehnung der KI-Nutzung wurde hingegen am deutlichsten abgelehnt. Die übrigen Maßnahmen, wie die Ergänzung, Reduktion oder Ersetzung klassischer Hausarbeiten, haben mittlere Werte.

#### **5.4.1.4 Alternative Prüfungsformate**

Bei den alternativen Prüfungsformaten bewerten die Befragten insbesondere praxisnahe Projektarbeiten als sinnvoll. Auch mündliche Prüfungen werden positiv eingeschätzt, während schriftliche Klausuren im Vergleich zurückhaltender beurteilt werden.

Aussage	MW	SD
Projektarbeiten	3,88	1,03
Mündliche Prüfungen	3,52	1,13
Klausuren	2,83	1,00

Tabelle 5.5: Bewertung alternativer Prüfungsformate

#### **5.4.2 Vertiefende statistische Analysen**

##### **5.4.2.1 Korrelationen und Multikollinearität**

Zur Untersuchung möglicher Zusammenhänge zwischen den erhobenen Variablen wurde eine Korrelationsmatrix erstellt.

Item		r
A	B	
Erfahrung KI	Haltung KI	0,53
Erfahrung KI	Häufigkeit KI	0,52
Ersetzung von Hausarbeiten	Erfahrung KI	-0,46
Erfahrung KI	KI verbessert Qualität & Arbeitsprozess	0,46
KI-Verbot	Kennzeichnung KI-Inhalte	0,44
Ersetzung von Hausarbeiten	Methodik	0,44
Ergänzung durch Alternativen	Methodik	0,44
Geschlecht	Haltung KI	-0,43
Geschlecht	Mündliche Prüfungen	-0,40
Reduktion von Hausarbeiten	Erfahrung KI	-0,40
Haltung KI	KI verbessert Qualität & Arbeitsprozess	0,40
Ergänzung durch Alternativen	Dokumentation der Prompts	0,36
Haltung KI	KI beeinträchtigt kognitive Leistung	-0,34
Mündliche Prüfungen	Haltung KI	0,33
Bedingte Zulassung KI	Abgabe von Logs	-0,33
Mündliche Prüfungen	Erfahrung KI	0,33
Reduktion von Hausarbeiten	Dokumentation der Prompts	0,32
KI-Verbot	Klausuren	0,32
Geschlecht	KI-Verbot	0,31
Geschlecht	Erfahrung KI	-0,3
Geschlecht	Methodik	-0,3

Tabelle 5.6: Übersicht relevanter Korrelationen ( $|r| \geq 0,30$ )

Dabei wurden Item-Kombinationen gewählt, bei denen auf Grundlage der Forschungsfragen ein inhaltlicher Zusammenhang der Auswirkungen von KI auf Hausarbeiten erwartet wurde. Als Schwellenwert für eine erhöhte Korrelation galt  $r = 0,30$ , da dieser Wert gemäß sozialwissenschaftlichen Konventionen als Beginn eines moderaten Zusammenhangs gilt. Insgesamt ergaben sich 22 bivariate Zusammenhänge, die diesen Wert überschritten haben und im weiteren Verlauf näher betrachtet wurden.

Ergänzend erfolgte eine Analyse der Varianzinflation (VIF), um potenzielle Multikollinearität zwischen den Prädiktorvariablen zu identifizieren. Dabei lag kein VIF-Wert oberhalb des Schwellenwerts von 5

vor. Somit bestehen keine Hinweise auf problematische Mehrfachabhängigkeiten innerhalb der berücksichtigten Merkmale.

#### 5.4.2.2 Gruppenunterschiede nach Subgruppen (Kruskal-Wallis, U-Test)

Auf Basis zuvor identifizierter Korrelationen wurden ausgewählte Items innerhalb verschiedener Subgruppen analysiert. Ziel war es, potenzielle Unterschiede in Abhängigkeit bestimmter Merkmale zu identifizieren.

Geschlecht				
Item	Kruskal-Wallis		Mann-Whitney-U	
	p	Signifikanz	p	Signifikanz
Mündliche Prüfungen	0,00	Ja	0,00	Ja
KI-Verbot	0,08	Nein	0,04	Ja
Erfahrung KI	0,04	Ja	0,06	Nein
Methodik	0,16	Nein	0,08	Nein
KI verbessert Qualität & Arbeitsprozess	0,11	Nein	0,32	Nein

Tabelle 5.7: Gruppenunterschiede nach Geschlecht (Kruskal-Wallis, Mann-Whitney-U)

Zur Analyse geschlechtsspezifischer Unterschiede wurden sowohl der Kruskal-Wallis-Test als auch ergänzend der Mann-Whitney-U-Test durchgeführt, da die dritte Kategorie „divers“ nur einfach besetzt war und somit keine belastbare Dreigruppenanalyse zuließ. Während sich bei der Einschätzung der eigenen KI-Erfahrung im Kruskal-Wallis-Test ein signifikanter Unterschied zeigte, wurde dieser im Mann-Whitney-U-Test nicht bestätigt.

Signifikante Unterschiede zwischen männlichen und weiblichen Teilnehmenden zeigten sich bei der Bewertung mündlicher Prüfungen und der Zustimmung zu einem generellen KI-Verbot. Während männliche Studierende mündliche Prüfungen positiver einschätzten, befürworteten weibliche Teilnehmende häufiger ein generelles KI-Verbot. Für die übrigen untersuchten Items ergaben sich keine signifikanten Unterschiede (alle  $p > 0,05$ ).

Erfahrungsabhängig zeigten sich signifikante Unterschiede bei Haltung, Nutzungshäufigkeit und der Einschätzung qualitativer Effekte.

Erfahrung		
Item	P	Signifikanz
KI verbessert Qualität und Arbeitsprozess	0,00	Ja
Haltung KI	0,04	Ja
Häufigkeit KI	0,04	Ja
Ersetzung von Hausarbeiten	0,13	Nein
Mündliche Prüfungen	0,69	Nein

Tabelle 5.8: Gruppenunterschiede nach Erfahrung (Kruskal-Wallis-Test)

Teilnehmende mit höherer KI-Erfahrung beurteilten deren Nutzen positiver und nutzten entsprechende Tools häufiger. Die Reduktion klassischer Hausarbeiten wurde hingegen häufiger von weniger erfahrenen Studierenden befürwortet. Keine signifikanten Unterschiede ergaben sich bei mündlichen Prüfungen und der vollständigen Ersetzung.

Haltung		
Item	P	Signifikanz
Erfahrung KI	0,00	Ja
KI verbessert Qualität und Arbeitsprozess	0,00	Ja
KI beeinträchtigt kognitive Leistung	0,01	Ja
Mündliche Prüfungen	0,01	Ja

Tabelle 5.9: Gruppenunterschiede nach Haltung (Kruskal-Wallis-Test)

Für das Merkmal Haltung zeigten sich bei allen betrachteten Items signifikante Unterschiede. Studierende mit positiver Einstellung gegenüber KI verfügten über mehr Erfahrung, bewerteten die Qualitätssteigerung durch KI höher und befürworteten mündliche Prüfungen häufiger. Teilnehmende mit kritischer Haltung äußerten hingegen stärkere Bedenken hinsichtlich potenziell negativer Effekte auf Denk- und Schreibprozesse.

Subgruppe	Item	p	Signifikanz
Bedingte Zulassung KI	Abgabe von Logs	0,03	Ja
KI-Verbot	Kennzeichnung KI-Inhalte	0,27	Nein
KI-Verbot	Klausuren	0,21	Nein
Reduktion von Hausarbeiten	Dokumentation der Prompts	0,05	Nein
Ergänzung durch Alternativen	Dokumentation der Prompts	0,07	Nein
Ergänzung durch Alternativen	Methodik	0,00	Ja
Ersetzung von Hausarbeiten	Methodik	0,65	Nein

Tabelle 5.10: Gruppenunterschiede nach Maßnahmen (Kruskal-Wallis-Test)

In der gruppenspezifischen Analyse ergaben sich signifikante Unterschiede bei zwei Items. Teilnehmende mit einer neutralen Haltung zur KI-Zulassung bewerteten die Abgabe von Screenshots oder Logs zustimmender als Befürwortende oder Ablehnende. Eine höhere Gewichtung des methodischen Aspektes wurde von jenen Studierenden gefordert, die alternative Prüfungsformate stärker befürworteten. Für alle weiteren Kombinationen aus Subgruppe und Maßnahme zeigten sich keine signifikanten Unterschiede.

#### 5.4.2.3 Regressionsanalyse

Zur Identifikation möglicher Einflussfaktoren auf die Bewertung ausgewählter Aussagen wurde eine ordinale Regressionsanalyse durchgeführt. Berücksichtigt wurden demografische sowie nutzungs- und haltungsbezogene Variablen.

Aussagen	Einflussfaktor	P	Signifikanz
KI mindert kognitive Leistung	Haltung	0,00	negativ
KI verbessert Qualität und Arbeitsprozess	Erfahrung	0,02	positiv
KI mindert Aussagekraft	Semester	0,04	negativ
KI erhöht Erwartungen	-	-	nein

Tabelle 5.11: Ergebnisse der Regressionsanalyse

In drei von vier Fällen zeigte sich jeweils ein signifikanter Zusammenhang, insbesondere mit Studienfortschritt, Erfahrung im Umgang mit KI und der persönlichen Haltung gegenüber der KI-Nutzung. Für die Aussage, dass KI die Erwartungen von Lehrenden erhöht, konnte kein signifikanter Einfluss festgestellt werden.

#### 5.4.2.4 Clusteranalyse

Zur Typisierung der Teilnehmenden im Umgang mit generativer KI wurde eine Clusteranalyse mit dem K-Means-Verfahren durchgeführt. Grundlage waren nutzungsbezogene Angaben und Bewertungen hochschulischer Maßnahmen. Die Analyse ergab drei Gruppen: Befürworter (28), Neutrale (14) und Skeptiker (24).

Aussage	Cluster/Gruppe					
	Befürworter		Neutral		Skeptiker	
	MW	SD	MW	SD	MW	SD
Erfahrung KI	6,29	0,85	4,29	1,27	5,17	1,09
Häufigkeit KI	4,57	0,5	3,5	0,65	3,92	0,72
Haltung KI	4,46	0,64	3,5	1,02	3,12	1,03
KI erhöht Erwartungen	3,82	1,25	3,64	1,01	3,33	0,82
KI mindert Aussagekraft	2,79	1,29	3,86	0,77	3,58	0,97
KI verbessert Qualität und Arbeitsprozess	4,82	0,39	4,07	0,62	3,71	1,23
KI mindert kognitive Leistung	2,79	1,37	3,29	1,07	3,67	1,13
KI-Verbot	1,18	0,39	1,14	0,36	2,54	1,02
KI-Zulassung (bedingt)	3,54	1,32	3,5	1,16	3,46	0,78
Ergänzung durch Alternativen	2,82	1,12	3,71	0,83	3,04	0,86
Reduktion von Hausarbeiten	2,29	1,15	4,43	0,51	2,92	0,93
Ersetzung von Hausarbeiten	1,61	0,92	4,07	1,0	2,88	1,12

Tabelle 5.12 Ergebnis der Clusteranalyse

Die Befürworter zeichnen sich durch hohe Werte bei Nutzungserfahrung und Zustimmung zum Nutzen von KI aus, verbunden mit einer überwiegend ablehnenden Haltung gegenüber Verboten und der vollständigen Ersetzung klassischer Hausarbeiten. Die neutrale Gruppe bewegt sich überwiegend im mittleren Bereich, zeigt jedoch vergleichsweise hohe Zustimmung zu alternativen Prüfungsformaten und deren vollständiger Substitution. Die skeptische Gruppe weist moderate Erfahrung, eine zurückhaltende Haltung gegenüber KI sowie höhere Zustimmungswerte zu restriktiven Maßnahmen und potenziellen Risiken auf.

## 5.5 Zwischen Integration und Abgrenzung

Die Ergebnisse der Befragung zeigen, dass Studierende den Einsatz generativer KI zunehmend als einen Teil der wissenschaftlichen Praxis ansehen. Besonders im Kontext von Hausarbeiten bewerten Studierende den Einsatz generativer KI-Tools überwiegend als unterstützend. Besonders stark fällt diese Einschätzung bei Personen mit höherer Nutzungserfahrung und positiver Haltung zur Technologie aus. Die statistisch signifikanten Zusammenhänge zwischen diesen Merkmalen und der Bewertung funktionaler Arbeitserleichterung deuten darauf hin, dass praktische Erfahrung die wahrgenommene Nützlichkeit maßgeblich beeinflusst. Auch die Regressionsanalyse bestätigt diesen Befund: Das Zielmerkmal „KI verbessert die Qualität und den Arbeitsprozess“ lässt sich signifikant durch Erfahrung erklären ( $p = 0,02$ ). Vergleichbare Ergebnisse finden sich auch in früheren Studien, etwa bei Bitkom (2023) oder der Frankfurt UAS (2024).

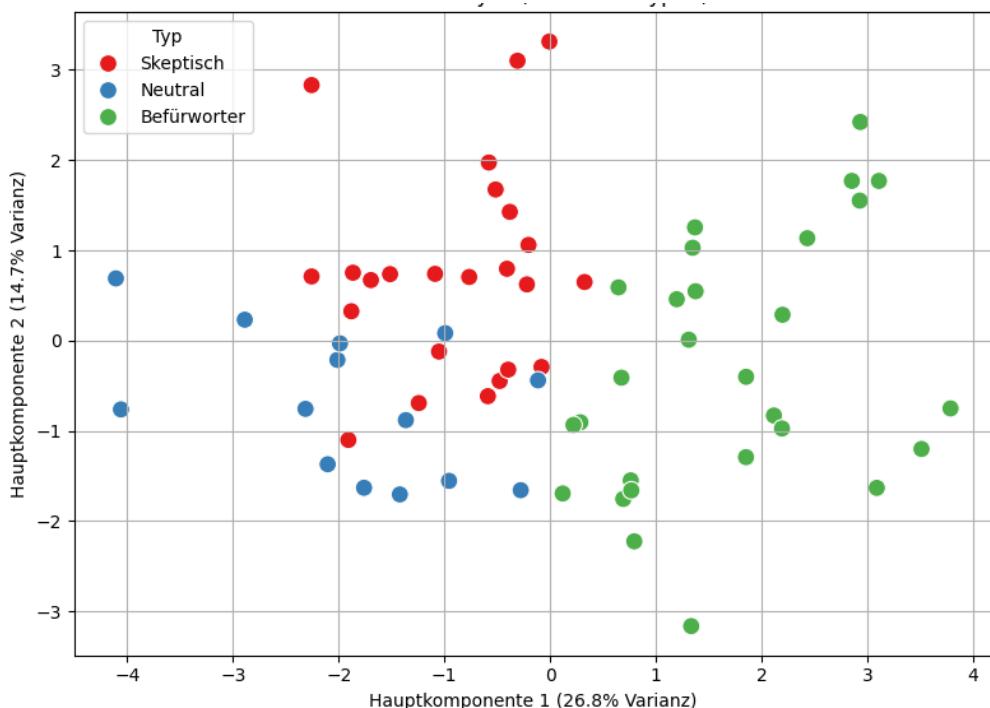


Abbildung 5.2: Visuelle Darstellung der Cluster

Neben den funktionalen Vorteilen treten auch kritische Perspektiven hervor. Besonders Teilnehmende mit negativer Haltung zur Technologie äußern häufiger Bedenken hinsichtlich möglicher Beeinträchtigungen kognitiver Leistungen. Der signifikante Zusammenhang mit der Haltung unterstreicht ein Spannungsfeld zwischen Effizienz und individueller Kompetenzentwicklung, das bereits im theoretischen Rahmen (Kap. 2.2) aufgegriffen wurde.

Im Rahmen der ersten Forschungsfrage wird deutlich, dass der Einsatz generativer KI-Tools zu einer Neubewertung klassischer Anforderungen führt. Studierende messen insbesondere dem methodischen Vorgehen und der kritischen Auseinandersetzung mit dem Thema eine höhere Relevanz bei. Ein möglicher Grund hierfür ist, dass diese Leistungen trotz Einsatz von KI überwiegend selbstständig erbracht werden müssen. Formale Kriterien wie sprachliche Gestaltung verlieren hingegen an Bedeutung, da sie vermutlich aus Sicht der Studierenden weitgehend von generativen KI-Tools übernommen werden können. Die Bewertung erfolgt dabei nicht unabhängig, sondern entlang der Frage, welche Anteile trotz KI-Nutzung noch als authentisch wahrgenommen werden.

Reflexive Formate wie begleitende Berichte oder Prompterläuterungen werden auch als wichtig erachtet. Das könnte daran liegen, dass dies Transparenz ermöglicht. Technikbasierte Nachweispflichten, wie KI-Erkennungs- oder Plagiatstools, stoßen dagegen auf Zurückhaltung. Eine mögliche Ursache liegt im mangelnden Vertrauen in deren Zuverlässigkeit und Fairness.

Diese Perspektive steht im Einklang mit der im Diskussionspapier der Hochschule RheinMain vertretenen Annahme, dass künftig nicht die Nutzung von KI, sondern die Prüfung verbleibender Eigenleistung im Vordergrund stehen sollte.

Diese Entwicklung spiegelt auch die im theoretischen Teil (Kapitel 2.4) beschriebene Unsicherheit darüber wider, wie die Bewertung schriftlicher Prüfungsleistungen künftig erfolgen soll, wenn menschliche und maschinelle Anteile zunehmend schwer zu trennen sind (University of Reading, 2024).

Die Bewertung hochschulischer Maßnahmen im Umgang mit generativer KI zeigt ebenfalls ein differenziertes Meinungsbild. Am meisten Zustimmung erfährt der Vorschlag, KI bedingt zuzulassen oder gezielt zu integrieren, auch wenn der Mittelwert insgesamt nur im neutralen Bereich liegt. Diese Präferenz lässt sich unter anderem dadurch erklären, dass viele Befragte eine grundsätzlich positive Haltung gegenüber KI einnehmen und deren Einfluss als förderlich einschätzen. Dagegen werden restriktive Maßnahmen wie ein generelles Verbot oder die vollständige Ersetzung klassischer Hausarbeiten mehrheitlich abgelehnt. Dies deutet darauf hin, dass Studierende KI nicht grundsätzlich als Bedrohung, sondern als integrierbare Unterstützung im Prüfungsprozess betrachten.

Die Clusteranalyse verdeutlicht, dass sich die Haltung zu institutionellen Regelungen stark entlang erfahrungs- und nutzungsbezogener Linien unterscheidet. Während Befürworter eine hohe Nutzung und positive Bewertungen aufweisen, äußern sich Skeptiker deutlich kritischer. Auffällig ist, dass gerade kritisch eingestellte Personen häufiger alternative Prüfungsformate befürworten. Befürworter hingegen lehnen alternative Formate häufiger ab und sprechen sich eher für die Beibehaltung klassischer Hausarbeiten aus. Dies lässt sich dahingehend interpretieren, dass sie den strategischen Nutzen von KI im bisherigen Prüfungsformat für sich erkannt haben. Ihre Bewertung scheint daher eher pragmatisch als didaktisch motiviert. Neutrale Personen neigen zu einer Offenheit gegenüber alternativen Formaten, ohne dabei klassische Strukturen grundsätzlich infrage zu stellen.

Ein Abgleich mit der Recherche zu hochschulischen Regelungen bestätigt diese Tendenzen: Die Mehrheit der untersuchten Hochschulen erlaubt den Einsatz generativer KI in Hausarbeiten grundsätzlich, knüpft diesen jedoch an Bedingungen wie Transparenz, Kennzeichnung und Eigenleistungsnachweis. Eine vollständige Ersetzung klassischer Prüfungsformate ist in keinem Fall vorgesehen. Damit decken sich die institutionellen Positionen weitgehend mit den Einschätzungen der Studierenden, die eine regulierte Integration gegenüber restriktiven Maßnahmen bevorzugen.

Im Kontext dieser Ergebnisse ist zu berücksichtigen, dass Erfahrung, Nutzung und Haltung zur KI im engen Zusammenhang stehen. Erfahrung entsteht meist durch wiederholte Nutzung, während eine positive Haltung auf erfolgreichen Anwendungen basiert. Umgekehrt kann eine kritische Haltung mit geringer Nutzung und entsprechend begrenzter Erfahrung einhergehen.

Künftige Studien sollten auf größeren und vielfältigeren Stichproben basieren, um die Ergebnisse zu überprüfen und besser auf verschiedene Kontexte übertragen zu können. Ergänzend wären qualitative Methoden hilfreich, um individuelle Einschätzungen und Beweggründe zur Nutzung von KI differenzierter zu erfassen. Langfristige Veränderungen in der Nutzung und Bewertung generativer KI im Hochschulkontext stellen ebenfalls ein relevantes Feld für weiterführende Untersuchungen dar.

## **5.6 Schlussfolgerungen und Implikationen für die Hochschulpraxis**

Die Studie ging der Frage nach, wie Studierende die Nutzung generativer KI im Kontext schriftlicher Prüfungen bewerten. Im Fokus standen dabei veränderte Anforderungen an Hausarbeiten sowie die Einschätzung hochschulischer Maßnahmen.

Die Ergebnisse zeigen, dass generative KI-Tools wie ChatGPT von 85% der Studierenden als funktionale Unterstützung im Schreibprozess wahrgenommen werden. Studierende mit höherer Erfahrung und positiver Haltung bewerten den Einsatz als effizient und hilfreich. Kritisch eingestellte Personen äußern dagegen Bedenken hinsichtlich möglicher Beeinträchtigungen der eigenen Denk- und Reflexionsleistungen. Insgesamt wird ein Spannungsverhältnis zwischen pragmatischer Nutzung und dem Anspruch auf individuelle Leistung deutlich.

Bezüglich der Anforderungen an Hausarbeiten werden inhaltliche und methodische Kompetenzen sowie die kritische Auseinandersetzung mit dem Thema als besonders relevant eingeschätzt. Formale Kriterien wie sprachliche Gestaltung treten demgegenüber in den Hintergrund. Die Bewertungen folgen einer funktionalen Logik: Entscheidend ist, welche Anforderungen trotz KI-Unterstützung noch als Ausdruck eigener Leistung gelten.

Auch die Bewertung hochschulischer Maßnahmen fällt differenziert aus. Pauschale Verbote werden überwiegend abgelehnt, während Konzepte zur kontrollierten Integration von KI breite Zustimmung finden. Alternative Prüfungsformate erhalten vor allem von kritisch eingestellten Personen Unterstützung, während erfahrene Nutzende stärker an klassischen Hausarbeiten festhalten. Die Einschätzungen institutioneller Regelungen hängen somit eng mit individuellen Erfahrungen und Haltungen zur KI zusammen.

Zusammenfassend lässt sich festhalten, dass Studierende den Einsatz generativer KI nicht als generelle Bedrohung, sondern als Herausforderung für bestehende Prüfungsformate begreifen. Sie fordern keine grundlegende Abschaffung, sondern eine Weiterentwicklung von Anforderungen und Bewertungskriterien, die den veränderten Rahmenbedingungen gerecht wird und Eigenleistung weiterhin sichtbar macht. Hochschulen sind damit aufgefordert, auf diese Entwicklungen konstruktiv zu reagieren und klare, didaktisch fundierte Regelungen zu etablieren, die sowohl technische Entwicklungen als auch pädagogische Zielsetzungen berücksichtigen.

## **5.7 Literaturverzeichnis**

Bitkom (2024): So digital sind Deutschlands Hochschulen. [online] Bitkom e.V. Verfügbar unter: <https://www.bitkom.org/Presse/Presseinformation/So-digital-sind-Deutschlands-Hochschulen> [Zugriff am: 27.05.2025].

BR (2023): Hausarbeiten mit ChatGPT – was bayerische Unis jetzt erlauben. [online] Bayerischer Rundfunk. Verfügbar unter: <https://www.br.de/nachrichten/netzwelt/hausarbeiten-mit-chatgpt-was-bayerische-unis-jetzt-erlauben,Ujl2Yze> [Zugriff am: 08.06.2025].

Chan, C. K. Y. & Hu, W. (2023): Students' voices on generative AI: perceptions, benefits, and challenges in higher education. The University of Hong Kong. [Zugriff am: 27.05.2025].

FernUniversität in Hagen (2023): KI-Leitfaden. Empfehlungen zum Umgang mit KI-Tools in Studium und Lehre. [online] Zentrum für Lernen und Innovation (ZLI), FernUniversität in Hagen. Verfügbar unter: [https://www.fernuni-hagen.de/zli/docs/6716\\_ki-leitfaden\\_-\\_din\\_a4\\_-\\_web.pdf](https://www.fernuni-hagen.de/zli/docs/6716_ki-leitfaden_-_din_a4_-_web.pdf) [Zugriff am: 08.06.2025].

- Frankfurt University of Applied Sciences (2024): Lehren, Lernen und Schreiben mit generativen KIAnwendungen: Empfehlungen und Impulse. Impulspapier (Stand: 11. März 2024). [online] Frankfurt University of Applied Sciences. Verfügbar unter: [https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren\\_Lernen\\_und\\_Schreiben\\_mit\\_generativen\\_KIAnwendungen\\_BeSt\\_BiB\\_Schreibwerkstatt\\_V4.2.pdf](https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren_Lernen_und_Schreiben_mit_generativen_KIAnwendungen_BeSt_BiB_Schreibwerkstatt_V4.2.pdf) [Zugriff am: 27.05.2025].
- Goethe-Universität Frankfurt (2024): Hinweise zur Nutzung von generativer KI in Studien- und Prüfungsleistungen. [online] Goethe-Universität Frankfurt am Main. Verfügbar unter: <https://www.goethe-university-frankfurt.de/132093840.pdf> [Zugriff am: 08.06.2025].
- HIS-Institut für Hochschulentwicklung (2025): KI in Studium und Lehre – Use-Cases und Gelingensbedingungen. HFD-Arbeitspapier Nr. 87. Berlin: Hochschulforum Digitalisierung. [Zugriff am: 30.05.2025]
- IU (2024): IU-Studie zeigt Potenzial von KI für gerechtere Notengebung. [online] IU Internationale Hochschule. Verfügbar unter: <https://www.iu.de/news/iu-studie-zeigt-potenzial-von-ki-fuer-gerechtere-notengebung/> [Zugriff am: 08.06.2025].
- IU (2023b): KI-Scanner an der Uni – Detektoren versagen. [online] IU Internationale Hochschule. Verfügbar unter: <https://www.iu.de/magazin/ki-scanner-an-der-uni-detektoren-versagen/> [Zugriff am: 08.06.2025].
- IU (2023a): Von ChatGPT bis DeepL – wie man KI-Tools beim Lernen richtig einsetzt. [online] IU Internationale Hochschule. Verfügbar unter: <https://www.iu.de/news/von-chatgpt-bis-deepl-wie-man-ki-tools-beim-lernen-richtig-einsetzt/> [Zugriff am: 08.06.2025].
- Johnston, H., Luo, C., Mallik, A. & Hamilton, J. (2024): Student perspectives on the use of generative artificial intelligence technologies in higher education. University of Sydney. [Zugriff am: 27.05.2025].
- LMU (2023b): Empfehlungen zum Umgang mit generativen und anderen KI-Tools. [online] Department of Statistics and Data Science, Ludwig-Maximilians-Universität München. Verfügbar unter: <https://cms-cdn.lmu.de/media/04-som/dss/downloads/empfehlungen-zum-umgang-mit-generativen-und-anderen-ki-tools.pdf> [Zugriff am: 08.06.2025].
- LMU (2023a): Rules for the use of AI tools in academic writing. [online] Department of Biology, Ludwig-Maximilians-Universität München. Verfügbar unter: <https://www.mcb.bio.lmu.de/downloads/rules-for-use-of-ai-tools.pdf> [Zugriff am: 08.06.2025].
- RWTH Aachen (2024c): Empfehlungen zum Umgang mit generativer KI in der wissenschaftlichen Praxis. [online] Betriebswirtschaftliches Forschungszentrum für Fragen der mittelständischen Wirtschaft (BFW), RWTH Aachen. Verfügbar unter: [https://www.bfw.rwth-aachen.de/global/show\\_document.asp?id=aaaaaaaaactfzlx&download=1](https://www.bfw.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaactfzlx&download=1) [Zugriff am: 08.06.2025].
- RWTH Aachen (2024b): Generative KI bei Abschlussarbeiten – Positionspapier des ICoM. [online] Institut für Unternehmenskybernetik (ICoM), RWTH Aachen. Verfügbar unter: <https://www.icom.rwth-aachen.de/cms/icom/das-institut/aktuelle-meldungen/~bjdpek/august-2024-generative-ki-bei-abschluss/> [Zugriff am: 08.06.2025].
- RWTH Aachen (2024a): Hinweise zur Nutzung von KI in wissenschaftlichen Arbeiten. [online] Center for Learning and Scientific Computing (CLS), RWTH Aachen. Verfügbar unter: [https://cls.rwth-aachen.de/global/show\\_document.asp?id=aaaaaaaaacocblkv](https://cls.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaacocblkv) [Zugriff am: 08.06.2025].
- Scribbr (2025): Die ChatGPT-Richtlinien der 100 größten deutschen Universitäten. [online] Scribbr.de. Verfügbar unter: <https://www.scribbr.de/ki-tools-nutzen/chatgpt-universitaere-richtlinien/> [Zugriff am: 29.05.2025].
- TUM (2024a): Formvorschriften für das Verfassen von Seminar- und Abschlussarbeiten im Fach Controlling. [online] Technische Universität München. Verfügbar unter: [https://www.fa.mgt.tum.de/fileadmin/w00chf/controlling/pdf/20240425\\_Formvorschriften\\_Controlling\\_1\\_.pdf](https://www.fa.mgt.tum.de/fileadmin/w00chf/controlling/pdf/20240425_Formvorschriften_Controlling_1_.pdf) [Zugriff am: 08.06.2025].

- TUM (2024b): Generative KI im Hochschulkontext – Herausforderungen und Empfehlungen. [online] Technische Universität München. Verfügbar unter: <https://mediatum.ub.tum.de/doc/1766632/1766632.pdf> [Zugriff am: 08.06.2025].
- Universität Hamburg (2024): Generative KI und Prüfungen. Informationen des Studienbüros Sozialwissenschaften. [online] Universität Hamburg. Verfügbar unter: <https://www.wiso.uni-hamburg.de/studienbuero-sozialwissenschaften/stichwortverzeichnis/gki-und-pruefungen.html> [Zugriff am: 08.06.2025].
- Universität Münster (2024): Umgang mit KI im Studium – Informationen für Studierende. [online] Universität Münster. Verfügbar unter: <https://www.uni-muenster.de/studium/orga/ki/index.html> [Zugriff am: 08.06.2025].
- Universität zu Köln (2023): FAQ zum Einsatz von KI-Tools wie ChatGPT in Prüfungsformaten. [online] Stabsstelle Recht. Verfügbar unter: <https://verwaltung.uni-koeln.de/stabsstelle02.1/content/faq/> [data/chatgpt/index\\_ger.html](https://data/chatgpt/index_ger.html) [Zugriff am: 08.06.2025].

## 6

## Vertrauen in Generative KI

*Quantitative Analyse der Validierung von ChatGPT-Ergebnissen durch Bachelorstudierende der Wirtschaftsinformatik an der Frankfurt University of Applied Sciences*

Julia Gimbel

### Abstract

Diese Arbeit untersucht das Vertrauen von Bachelorstudierenden der Wirtschaftsinformatik an der Frankfurt University of Applied Sciences (FRANKFURT UAS) in die von ChatGPT generierten Informationen.

In einer Zeit, in der Generative Künstliche Intelligenz (GenKI) zunehmend in akademischen Kontexten eingesetzt wird, ist es von entscheidender Bedeutung, die Wahrnehmungen und Einstellungen der Studierenden zu verstehen. Diese quantitative Analyse basiert auf einer Umfrage unter 66 Studierenden zu deren Vertrauen in ChatGPT.

Die Ergebnisse zeigen, dass das Vertrauen in die von ChatGPT gelieferten Informationen stark von der persönlichen Erfahrung und der kritischen Auseinandersetzung mit der Technologie abhängt. Studierende mit umfangreicherer Erfahrung neigen dazu, den generierten Inhalten mehr Vertrauen zu schenken. Zudem wird die Bedeutung der Validität der Informationen als zentraler Faktor für die Akzeptanz von GenKI hervorgehoben.

Die Arbeit schließt mit Empfehlungen für Hochschulen, gezielte Schulungsangebote zu entwickeln, um den verantwortungsvollen Einsatz von GenKI in der Lehre zu fördern, und bietet Ausblicke auf zukünftige Forschungsmöglichkeiten im Bereich der digitalen Bildung.

**Schlagwörter:** Generative KI, Vertrauen in Technologie, ChatGPT, Hochschulbildung, Validität von Informationen, Studierendenwahrnehmung, Umfrageforschung, Plagiatsangst, Validierungsabsicht

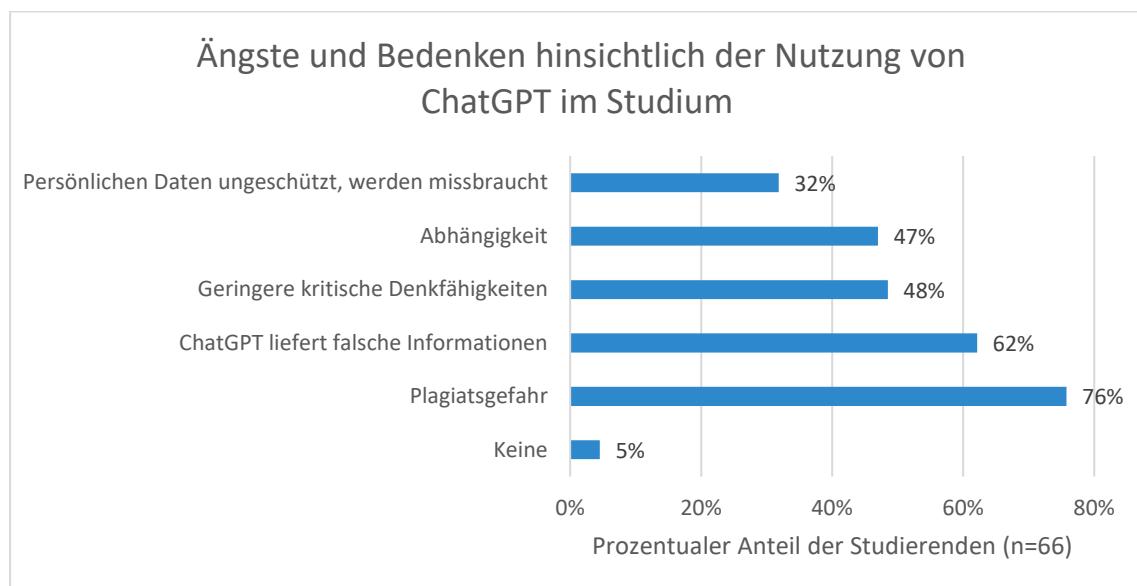


Abbildung 6.1: Ängste und Bedenken zur Nutzung von ChatGPT im Studium

## **6.1 Einleitung**

In der heutigen digitalen Ära nimmt die Nutzung von Künstlicher Intelligenz (KI) in diversen Lebensbereichen stetig zu. Besonders im akademischen Kontext gewinnt Generative Künstliche Intelligenz (GenKI) an Bedeutung, da sie Studierenden neue Möglichkeiten zur Informationsbeschaffung und -verarbeitung bietet. Diese Arbeit beschäftigt sich mit dem Vertrauen von Bachelorstudierenden der Wirtschaftsinformatik an der Frankfurt University of Applied Sciences in die von ChatGPT generierten Ergebnisse. Die Relevanz des Themas ergibt sich aus der zunehmenden Integration von KI-Tools in den Bildungsbereich.

Die Motivation für diese Arbeit liegt in der Beobachtung, dass trotz der wachsenden Verbreitung von GenKI in der Hochschulbildung Unsicherheiten hinsichtlich der Validität der generierten Informationen bestehen. Das Vertrauen der Studierenden in diese Technologien ist entscheidend für deren erfolgreiche Anwendung. Daher ist es wichtig, die Wahrnehmungen und Einstellungen der Studierenden zu untersuchen, um ein besseres Verständnis für die Akzeptanz von GenKI in der akademischen Welt zu entwickeln.

Das Ziel dieser Arbeit ist es, das Vertrauen von Studierenden in ChatGPT quantitativ zu analysieren und Faktoren zu identifizieren, die dieses Vertrauen beeinflussen.

Die zentrale Forschungsfrage dieser Arbeit lautet: Inwiefern vertrauen Bachelorstudierende der Wirtschaftsinformatik an der Frankfurt UAS den von ChatGPT generierten Informationen? Darüber hinaus werden folgende Unterfragen behandelt, welche unter anderem den Rahmen für die Analyse und Diskussion der Ergebnisse bilden: Welche Faktoren beeinflussen das Vertrauen in die von ChatGPT gelieferten Inhalte? Wie wichtig ist die Validität dieser Informationen für die Studierenden? Inwiefern validieren Studierende die generierten Informationen von ChatGPT?

## **6.2 Definitionen und Theorien**

Im Rahmen dieses Kapitels werden zentrale Begriffe und theoretische Konzepte definiert, die für das Verständnis des Vertrauens in Generative Künstliche Intelligenz und deren Anwendung im Hochschulkontext von Bedeutung sind. Zunächst wird der Begriff der Generativen KI erläutert, gefolgt von einer detaillierten Betrachtung von ChatGPT als prominentes Beispiel. Anschließend wird die Validität von Informationen, die von solchen KI-Systemen bereitgestellt werden, untersucht, da sie eine entscheidende Rolle für das Vertrauen der Nutzer spielt. In diesem Zusammenhang wird auch auf bestehende Studien in diesem Feld eingegangen, aus denen Hypothesen abgeleitet werden, die es zu prüfen gilt.

### **6.2.1 Validität einer Information**

Die Validität einer Information beschreibt, inwieweit eine Information inhaltlich korrekt, wahrheitsgemäß, logisch konsistent, kontextangemessen und relevant ist (Metzger, 2007, S. 2080).

Im Kontext der Nutzung von KI-gestützten Systemen wie ChatGPT umfasst Validität sowohl die Faktentreue der generierten Inhalte als auch ihre thematische Passgenauigkeit in Bezug auf die jeweilige Aufgabenstellung (Metzger, 2007, S. 2079).

Eine valide Information erfüllt demnach folgende Kriterien:

- Factual Accuracy: Die Information stimmt mit objektiv überprüfbar Fakten oder anerkannten wissenschaftlichen Erkenntnissen überein (Metzger, 2007, S. 2081).

- Logical Consistency: Die Information ist innerhalb ihrer Argumentationsstruktur widerspruchsfrei (Flanagin & Metzger, 2008, S. 6).
- Contextual Appropriateness: Die Information berücksichtigt den inhaltlichen und situativen Kontext korrekt (Nemt-Allah et al., 2024, S.5).
- Relevance: Die Information trägt sinnvoll und direkt zur Lösung der gestellten Frage oder Aufgabe bei (Abdaljaleel et al., 2024, S.2).

Gerade bei der Nutzung von GenKI-Modellen wie ChatGPT ist die Validitätsprüfung entscheidend, da die KI zwar plausible, aber nicht notwendigerweise korrekte Antworten generiert (Abdaljaleel et al., 2024, S.2).

### 6.2.2 Vertrauen bei der Nutzung von ChatGPT

Im Kontext der Nutzung von KI-gestützten Systemen wie ChatGPT bezeichnet Vertrauen das Ausmaß, in dem Nutzende darauf vertrauen, dass die Technologie verlässlich, akkurat, transparent und sicher arbeitet (McKnight et al., 2002, S. 340). Vertrauen umfasst sowohl die Erwartung, dass die von ChatGPT generierten Inhalte faktisch korrekt und verständlich sind, als auch die Überzeugung, dass die Nutzung der Technologie keine negativen Konsequenzen wie Datenmissbrauch oder Fehlinformationen nach sich zieht (Abdaljaleel et al., 2024, S.2).

Vertrauen in KI-Technologien wie ChatGPT basiert auf drei zentralen Dimensionen:

1. Kompetenzvertrauen: Die Überzeugung, dass ChatGPT über die technischen Fähigkeiten verfügt, Aufgaben korrekt, effizient und erwartungsgemäß zu erfüllen (McKnight et al., 2002, S. 337).
2. Integritätsvertrauen: Die Erwartung, dass ChatGPT "richtig" arbeitet, also keine absichtlichen Täuschungen oder Verzerrungen produziert und Inhalte fair wiedergibt (McKnight et al., 2002, S. 337).
3. Benevolenzvertrauen: Das Vertrauen darauf, dass die Technologie im Interesse des Nutzenden handelt und nicht zu Schaden oder Nachteilen führt, etwa durch unsichere Datenverarbeitung oder fehlerhafte Informationen (McKnight et al., 2002, S. 337).

### 6.2.3 Schwerpunkt: Validität von Informationen

Die Validität der von ChatGPT gelieferten Informationen nimmt im Rahmen des in Abbildung 2 betrachteten Modells eine zentrale Rolle ein, insbesondere auf der Ebene der verhaltensbezogenen Komponente, konkret innerhalb der Phase der Intention zur Überprüfung von Informationen (intention to verify information) (Menon & Shilpa, 2023, S. 2ff.).

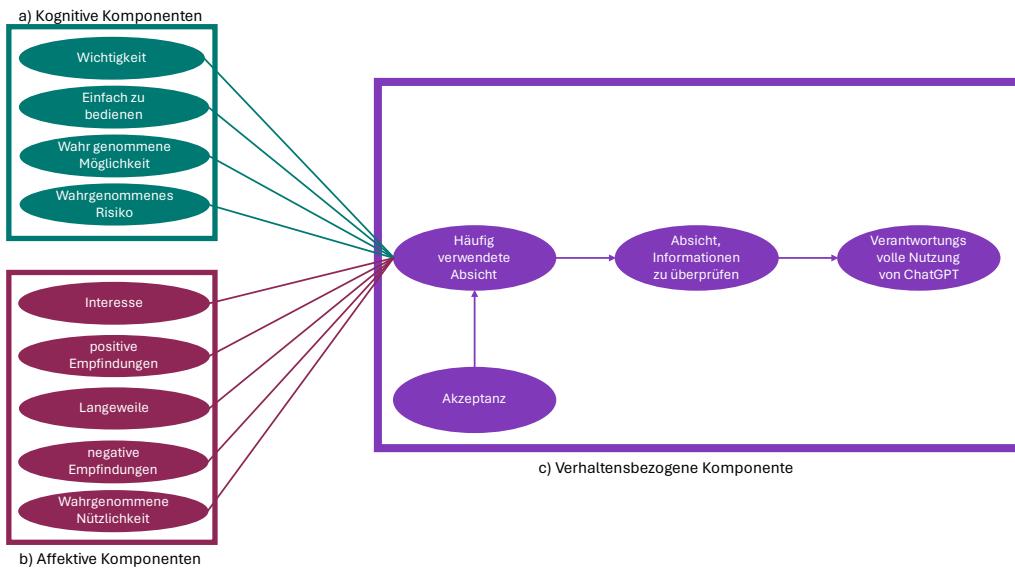


Abbildung 6.2: Komponentenübersicht

Studierende, die ChatGPT regelmäßig verwenden, entwickeln im Laufe ihrer Nutzungserfahrung eine bewusste Intention, die durch die KI generierten Inhalte kritisch zu überprüfen, anstatt diese unreflektiert zu übernehmen (Lund & Wang, 2023, S. 27f.).

Generative KI-Modelle wie ChatGPT sind prinzipiell in der Lage, fehlerhafte, unvollständige oder verzerrte Informationen zu erzeugen, da sie auf Wahrscheinlichkeitsmodellen basieren und keinen echten Wahrheitsbegriff kennen (OpenAI, 2022).

Ohne eine bewusste Überprüfung der gelieferten Inhalte besteht die Gefahr, dass unzuverlässige Informationen ungefiltert in akademische Arbeiten oder Entscheidungsprozesse übernommen werden (Lund & Wang, 2023, S. 27f.).

Die Validität der Informationen beeinflusst daher nicht nur das Vertrauen der Nutzenden in ChatGPT, sondern wirkt sich unmittelbar auf deren Verhaltensabsichten aus (Slade et al., 2015, S. 867).

Studierende, die die Qualität und Zuverlässigkeit der Inhalte aktiv hinterfragen und kritisch evaluieren, sind eher in der Lage, ChatGPT reflektiert, gezielt und verantwortungsvoll in ihre akademischen Lern- und Arbeitsprozesse zu integrieren (Menon & Shilpa, 2023, S. 11f.).

Die aktive Validierungsabsicht wirkt hier als eine Art Schutzmechanismus, der verhindert, dass fehlerhafte Informationen unhinterfragt übernommen werden, und der zugleich die langfristige und nachhaltige Integration von ChatGPT in die Hochschulbildung ermöglicht (Mhlanga, 2023, S. 16).

#### 6.2.4 Überblick über bestehende Forschung zu Generativer KI in der Bildung

Die Forschung zur Nutzung von Generativer Künstlicher Intelligenz (GenKI) in der Bildung ist vielfältig und zeigt unterschiedliche Perspektiven auf das Vertrauen von Studierenden in solche Technologien. Eine Umfrage, die an zwei großen Universitäten in den USA und Indien durchgeführt wurde, befasste sich mit der Nutzung von GenKI-Tools wie ChatGPT im Kontext der Informatikausbildung. Die Ergebnisse zeigten, dass eine Mehrheit der 253 Befragten bereits Erfahrung mit GenKI hatte, wobei 47 % Vertrauen in diese Technologien äußerten. Interessanterweise gab es eine signifikante

Korrelation zwischen dem Vertrauen in GenKI und der Selbstwahrnehmung von Motivation und Selbstvertrauen in Programmierfähigkeiten, insbesondere bei Erstgeneration-Studierenden. Dennoch äußerten einige Teilnehmer Bedenken hinsichtlich Plagiate und der Entwicklung schlechter Gewohnheiten im Umgang mit GenKI (Amoozadeh, et al., 2024, S. 1f.).

Eine weitere Studie, die sich mit den Einstellungen von 499 Studierenden an verschiedenen Universitäten in Peru gegenüber ChatGPT in ihren akademischen Aktivitäten befasste, nutzte das Technology Acceptance Model (TAM) als theoretischen Rahmen. Über 75 % der Studierenden gaben an, ChatGPT verwendet zu haben. Die Studie identifizierte positive Korrelationen zwischen dem Vertrauen in ChatGPT und der Bereitschaft, die Technologie häufig zu nutzen sowie Informationen zu verifizieren. Diese Ergebnisse unterstreichen die Bedeutung von Vertrauen und verantwortungsvollem Umgang mit GenKI in akademischem Kontext. Studierende, die bereits Erfahrungen mit digitalen Technologien wie Lernmanagementsystemen, Online-Datenbanken oder E-Learning-Tools gemacht haben, sind tendenziell besser in der Lage, die Qualität und Zuverlässigkeit der von GenKI generierten Inhalte zu bewerten. Dies zeigt, dass die Ausbildung in der Nutzung digitaler Technologien eine wichtige Rolle dabei spielt, wie Studierende GenKI im Bildungsbereich wahrnehmen und anwenden (Acosta-Enriquez et al., 2024, S. 7ff.).

### **6.2.5 Vertrauensfragen in der Nutzung von KI**

Das Vertrauen in Künstliche Intelligenz (KI) spielt eine entscheidende Rolle in der Akzeptanz und Nutzung von Technologien wie ChatGPT im Bildungsbereich. Eine Untersuchung, die an der University of Hong Kong durchgeführt wurde, ergab, dass über 75 % der 399 befragten Studierenden bereits Erfahrungen mit ChatGPT gemacht hatten. Die Studie identifizierte wesentliche Faktoren, die das Vertrauen der Studierenden in GenAI beeinflussen, darunter die wahrgenommene Nützlichkeit und Benutzerfreundlichkeit. Studierende, die die Technologie kritisch analysiert hatten, bewerteten die Praktikabilität des Tools positiver. Dies legt nahe, dass ein tiefes Verständnis der Technologie und ihrer Funktionsweise das Vertrauen der Nutzer stärken kann (Hu & Chan, 2023, S.5ff.).

In einer anderen Studie, die 500 Universitätsstudenten in den Philippinen befragte, wurde das Vertrauen in ChatGPT ebenfalls thematisiert. Die Ergebnisse zeigten, dass das Vertrauen in die Technologie eng mit der Absicht verknüpft ist, diese regelmäßig zu nutzen. Studierende, die ein hohes Maß an Vertrauen in die Technologie hatten, waren eher bereit, ChatGPT als nützlich und effektiv in ihren akademischen Aktivitäten zu betrachten. Die Bedenken hinsichtlich der Qualität der generierten Inhalte und der möglichen Konsequenzen einer unkritischen Nutzung bleiben jedoch bestehen (Obenza et al., 2023, S. 8).

### **6.2.6 Vergleich mit anderen Technologien**

Der Vergleich von Generativer Künstlicher Intelligenz (GenKI) mit anderen Technologien im Bildungsbereich zeigt sowohl Gemeinsamkeiten als auch Unterschiede in der Akzeptanz und Nutzung. Während viele Studierende bereits mit Technologien wie Lernmanagementsystemen oder Online-Datenbanken vertraut sind, bringt die Nutzung von GenKI wie ChatGPT spezifische Herausforderungen mit sich. Eine Studie über die Nutzung von ChatGPT in akademischen Kontexten hat gezeigt, dass Studierende oft Schwierigkeiten haben, die Qualität und Validität der von der KI generierten Inhalte zu bewerten. Dies steht im Gegensatz zu traditionelleren Technologien, bei denen die Quellen und Informationen in der Regel klarer identifizierbar sind (Obenza et al., 2023, S. 8).

Ein weiterer Aspekt ist die Interaktivität von GenKI. Im Vergleich zu statischen Lernressourcen ermöglicht GenKI eine dynamische Interaktion, die den Lernprozess anpassen kann. Die Umfrage

unter Studierenden an der University of Hong Kong zeigte, dass die Benutzerfreundlichkeit und die wahrgenommene Nützlichkeit von GenKI entscheidend für die Akzeptanz sind (Hu & Chan, 2023, S. 8). Diese Faktoren sind auch bei anderen Technologien relevant, jedoch ist die subjektive Wahrnehmung der Risiken und Herausforderungen bei der Nutzung von GenKI oft ausgeprägter. Insbesondere Bedenken hinsichtlich Plagiate und der Erzeugung von ungenauen Informationen stellen eine zusätzliche Hürde dar, die bei der Nutzung traditioneller Technologien weniger stark ausgeprägt ist (Amoozadeh, et al., 2024, S. 2f.).

### 6.2.7 Abgeleitete Hypothesen

Aus den zuvor explorativ beschriebenen Studien werden vier Hypothesen abgeleitet, welche in der vorliegenden Arbeit untersucht, beziehungsweise widerlegt oder bestätigt werden.

1. Hypothese: Es besteht ein positiver Zusammenhang zwischen der Selbstwahrnehmung von Motivation und Selbstvertrauen in Programmierfähigkeiten und dem Vertrauen in Generative KI bei Studierenden. Die Studie von Amoozadeh et al. (2024) zeigt, dass Erstgeneration-Studierende, die mehr Vertrauen in Generative KI haben, auch ein höheres Selbstvertrauen in ihren Programmierfähigkeiten aufweisen (Amoozadeh, et al., 2024, S. 6).
2. Hypothese: Studierende, die bereits Erfahrungen mit digitalen Technologien haben, weisen ein höheres Vertrauen gegenüber der Nutzung von KI-Tools wie ChatGPT auf. Laut der Studie von Acosta-Enriquez et al. (2024) sind Studierende mit mehr Erfahrung in digitalen Technologien besser in der Lage, die Qualität und Zuverlässigkeit von Generativen KI-Inhalten zu bewerten (Acosta-Enriquez et al., 2024, S. 13ff.).
3. Hypothese: Bedenken hinsichtlich Plagiate und der Qualität der von Generativer KI generierten Inhalte haben einen negativen Einfluss auf das Vertrauen in diese Technologie. Die Studie von Obenza et al. (2023) zeigt, dass trotz positiver Erfahrungen auch Bedenken bestehen, die das Vertrauen in Generative KI beeinträchtigen (Obenza et al., 2023, S. 11).
4. Hypothese: Studierende, die Generative KI kritisch analysieren und hinterfragen, zeigen ein höheres Maß an Vertrauen in die Technologie im Vergleich zu Studierenden, die dies nicht tun. Die kritische Auseinandersetzung mit der Technologie kann das Vertrauen und die Akzeptanz fördern, wie in der Studie von Hu & Chan (2023) angedeutet wird (Hu & Chan, 2023, S. 14f.).

## 6.3 Methodik

Im nachfolgenden Kapitel wird aufgezeigt, wie die vorliegende Forschungsfrage untersucht wurde. Dazu wurde sowohl eine intensive Literaturrecherche als auch eine Umfrage durchgeführt. Nachfolgend wird die Zielgruppe der Umfrage definiert und die Genauigkeit wird ermittelt. Die Repräsentativität der Stichprobe ist mit einem 95-prozentigem Konfidenzintervall gegeben.

### 6.3.1 Forschungsdesign und -ansatz

Der theoretische Hintergrund wurde durch eine eingehende Literaturrecherche untersucht. Der Auswahlprozess erfolgte in zwei Schritten: Zunächst wurden die Arbeiten anhand ihrer Titel, Zusammenfassungen und Schlüsselwörter gesichtet. Im zweiten Schritt wurden die Arbeiten ausgewählt, die nach der Durchsicht, der im ersten Schritt identifizierten Arbeiten sowie deren Volltext die relevanten Informationen enthielten.

Der Forschungsansatz dieser wissenschaftlichen Arbeit ist sowohl explorativ als auch korrelativ. Demnach werden sowohl neue Einsichten gewonnen als auch Beziehungen zwischen zwei oder mehreren Variablen untersucht, ohne Einflussnahme.

### 6.3.2 Datenerhebung

Zur Errechnung des mathematischen Errors beziehungsweise des Präzisionslevels wird folgende Formel verwendet (Bortz & Schuster, 2010, S.235):  $n = \left( \frac{1,96 s}{\Delta} \right)^2$

Diese Formel wird entsprechend nach Delta umgestellt und 1,0640 wird für die Standardabweichung eingesetzt, da die Daten der Referenzfrage „Wie stehen Sie zur Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten?“ verwendet werden. „n“ steht innerhalb der Formel für die Stichprobengröße, weshalb hier 66 eingesetzt wird. Der 1,96-Wert entspricht einem 95-prozentigen Konfidenzintervall (Schuster & Liesen, 2017, S. 117). „ $\Delta$ “ gibt den mathematischen Error beziehungsweise das Präzisionslevel an.

$$\Delta = \sqrt{\frac{1,96 * 1,0640}{66}} = 0,17776$$

Das Ergebnis von 17,78 Prozent ist demnach die Error Rate, welche beschreibt, zu wie viel Prozent die Ergebnisse der Stichprobe im positiven-, sowie im negativen von der Grundgesamtheit abweichen können.

### 6.3.3 Datenanalyseverfahren

Zur Auswertung der Umfragedaten werden unter Zuhilfenahme der Applikation Minitab und Microsoft Excel eine multiple Regressionsanalyse durchgeführt sowie ein Kruskal-Wallis-Test.

Die multiple Regressionsanalyse, untersucht den Einfluss von mehreren unabhängigen Variablen auf eine abhängige Variable. Sie bestimmt, wie exakt die unabhängigen Variablen die abhängige Variable vorhersagen können und welche Variablen den größten Einfluss haben. Hierzu wird zunächst eine Gleichung aufgestellt, die die Beziehung zwischen der abhängigen Variablen (z. B. Vertrauen in KI) und mehreren unabhängigen Variablen beschreibt. Die Methode ermöglicht es, den Einfluss jeder unabhängigen Variablen zu quantifizieren, während die anderen Variablen konstant gehalten werden (Multiple Regressionsanalyse, o. D.-b.).

Der Kruskal-Wallis-Test wird zur Untersuchung von Unterschieden zwischen drei oder mehr unabhängigen Gruppen verwendet. Er wird bei nicht normalverteilten Daten verwendet und überprüft, ob es signifikante Unterschiede zwischen den Medianen der Gruppen gibt. Er ist besonders nützlich, wenn die Daten ordinal sind, also eine bestimmte Reihenfolge oder Rangordnung besitzen, was in der Vorliegenden Arbeit für den Großteil der Variablen der Fall ist (McClenaghan, 2024).

Mithilfe von Excel wurden zunächst die Daten aufbereitet und umkodiert, sodass ordinale Antwortmöglichkeiten, wie beispielsweise „sehr negativ“, „negativ“, „neutral“, „positiv“ und „sehr positiv“ in ihre natürliche Reihenfolge 1-5 vorkommen. Somit lassen sich die quantitativen Werte der Antworten statistisch auswerten.

Da die Umfrage für jedes Antwortfeld eine Eingabe erzwingt, existieren keine fehlenden Werte. Auch Ausreißer wurden im Zusammenhang mit dieser Forschung nicht identifiziert.

## 6.4 Auswertung

In diesem Kapitel werden die Ergebnisse der Analyse präsentiert. Zunächst werden einige deskriptive Statistiken mittels Microsoft Excel generiert. Anschließend werden Korrelationen zwischen verschiedenen Variablen mittels Excel untersucht, welche das Vertrauen in Generative KI beeinflussen könnten. Diese umfassen beispielsweise die Erfahrung mit ChatGPT, die Nutzungshäufigkeit oder auch die wahrgenommene Nützlichkeit. Diese Korrelationen zeigen, wie stark die Beziehungen zwischen den verschiedenen Variablen sind. Anschließend werden die zuvor aufgestellten Hypothesen aus Kapitel 2.9 bestätigt oder widerlegt und eine multiple Regressionsanalyse mithilfe von Minitab wird durchgeführt, um die Faktoren zu identifizieren, die das Vertrauen in Generative KI am stärksten beeinflussen.

### 6.4.1 Deskriptive Statistiken

Abbildung 3 zeigt, wie die Studierenden die Vertrauenswürdigkeit, der von ChatGPT bereitgestellten Informationen für ihre akademischen Arbeiten ansehen.

Es ist zu erkennen, dass die Mehrheit der Studierende die Informationen als „eher nicht vertrauenswürdig“ mit 30% bis „neutral vertrauenswürdig“ mit 35% einstufen. Als vertrauenswürdig empfinden 27% der Befragten die von ChatGPT generierten Informationen. Lediglich eine Person empfindet die von ChatGPT generierten Informationen „sehr vertrauenswürdig“.

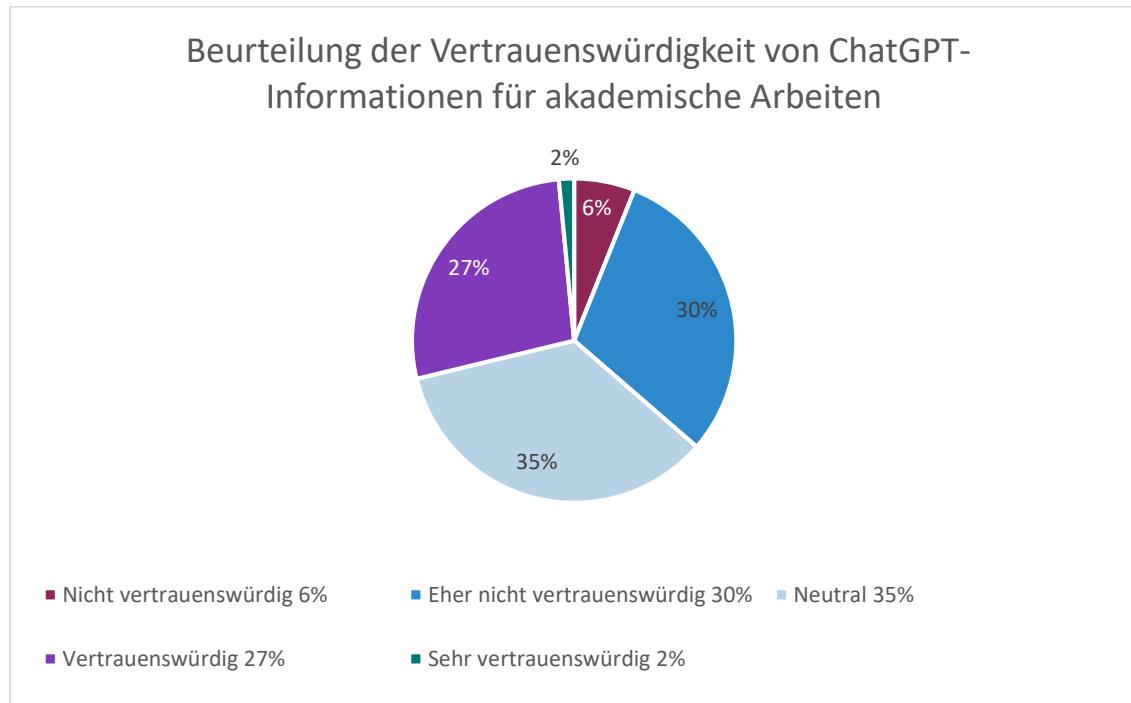


Abbildung 6.3: Beurteilung der Vertrauenswürdigkeit von ChatGPT-Informationen für akademische Arbeiten

Die Validität, der von ChatGPT generierten Informationen für akademische Arbeiten ist 38% der Studierenden wichtig, 44% der Befragten sogar sehr wichtig. Folglich ist es den Studierenden wichtig die von ChatGPT generierten Informationen zu validieren. Abbildung 4 schlüsselt die Wichtigkeit der Validität, der von ChatGPT bereitgestellten Informationen für akademische Arbeiten auf.

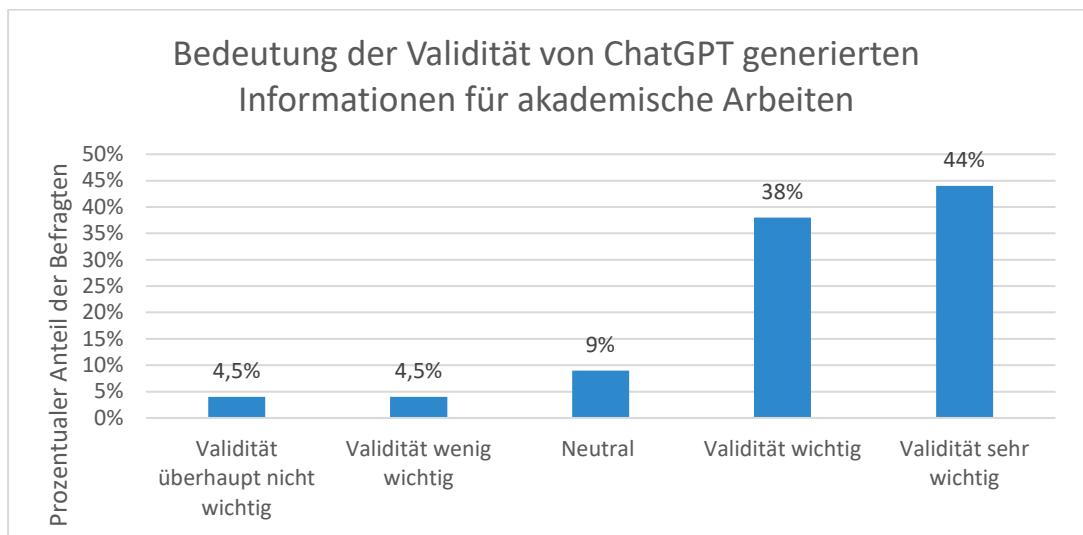


Abbildung 6.4: Bedeutung der Validität von ChatGPT generierten Informationen für akademische Arbeiten

Abbildung 5 zeigt, wie viel Zeit die befragten Studierenden für die Validierung, der von ChatGPT bereitgestellten Informationen, verwenden. Ein Großteil der Befragten verwendet dabei weniger als 30 Minuten je Ergebnis um die Informationen von ChatGPT zu validieren. Einige nehmen sich auch bis zu 1 Stunde Zeit, die wenigsten jedoch länger. Im Durchschnitt verwenden die Studierenden 28 Minuten für die Validierung der Ergebnisse.

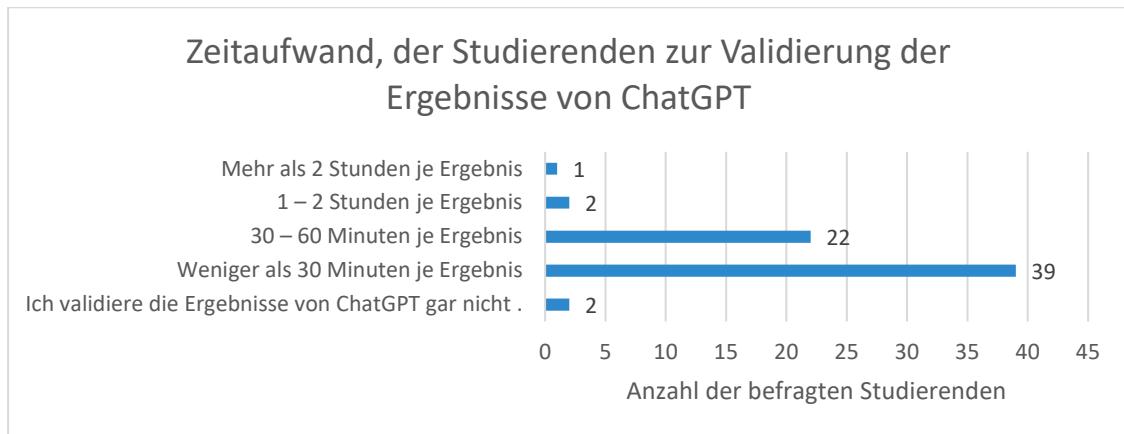


Abbildung 6.5: Zeitaufwand, der Studierenden zur Validierung der Ergebnisse von ChatGPT

Ängste und Bedenken hinsichtlich der Nutzung von ChatGPT im Studium wurden innerhalb der Umfrage ebenfalls abgefragt. Hierbei wurden bereits sechs Antwortoptionen vorgegeben, welche die Studierenden auswählen konnten. Dabei mussten Sie sich nicht für eine Antwortoption entscheiden, sondern konnten alle Antwortmöglichkeiten auswählen, welche auf Sie zutreffend erschienen. Dabei ist auffällig, dass bis auf 5% der Befragten, Bedenken oder Ängste hinsichtlich der Nutzung von ChatGPT im Studium haben. Die meisten haben Angst vor einem Plagiatsvorwurf. In Abbildung 6 ist ersichtlich, dass über die Hälfte der Befragten Bedenken hat, dass ChatGPT falsche

Informationen liefert. Fast die Hälfte aller Studierenden fühlt sich bereits von ChatGPT abhängig und bemerkt, dass ihre kritische Denkfähigkeit abnimmt.

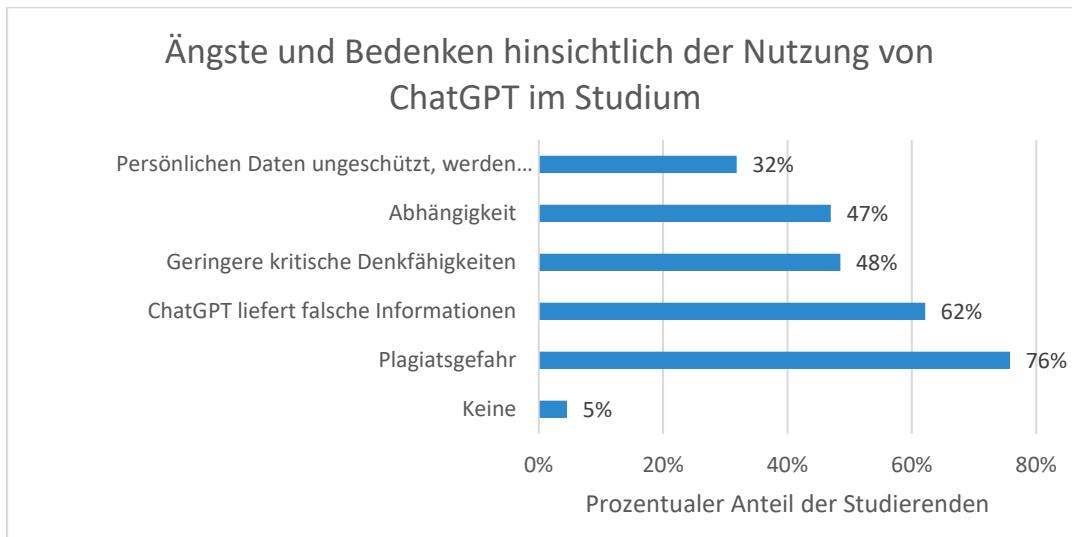


Abbildung 6.6: Ängste und Bedenken hinsichtlich der Nutzung von ChatGPT im Studium

Mittels Minitab wurde eine grundlegende statistisch Auswertung zur Verteilung der Daten, sowie zur Identifikation des Medians und der Standardabweichung erstellt. Diese ist in Abbildung 7 sichtbar und zeigt, dass die Daten der Umfrage nicht normalverteilt sind. Dies wird sichtbar durch das Histogramm, welches die Datenverteilung visuell zeigt. Die rote Kurve innerhalb des Histogramms stellt eine Normalverteilung der Daten dar, die hier vorliegenden Daten passen jedoch nicht gut zu dieser roten Kurve, was darauf hindeutet, dass die Daten der Umfrage nicht normalverteilt sind.

Auch der P-Wert des Anderson-Darlings Normalitätstests ist kleiner als 0,05. Dies weist darauf hin, dass die Nullhypothese der Normalverteilung abzulehnen ist. Das bedeutet, dass die Daten nicht normalverteilt sind.

Der Median der Nutzung von ChatGPT beim Verfassen von wissenschaftlichen Hausarbeiten liegt bei 4. Dies bedeutet die Befragten nutzen ChatGPT im universitären Kontext, also beispielsweise bei Hausarbeiten oder Klausurvorbereitung „häufig“. Da der Median robuster gegenüber Ausreißern ist als der Mittelwert (Mean), wird nicht weiter auf den Mean eingegangen.

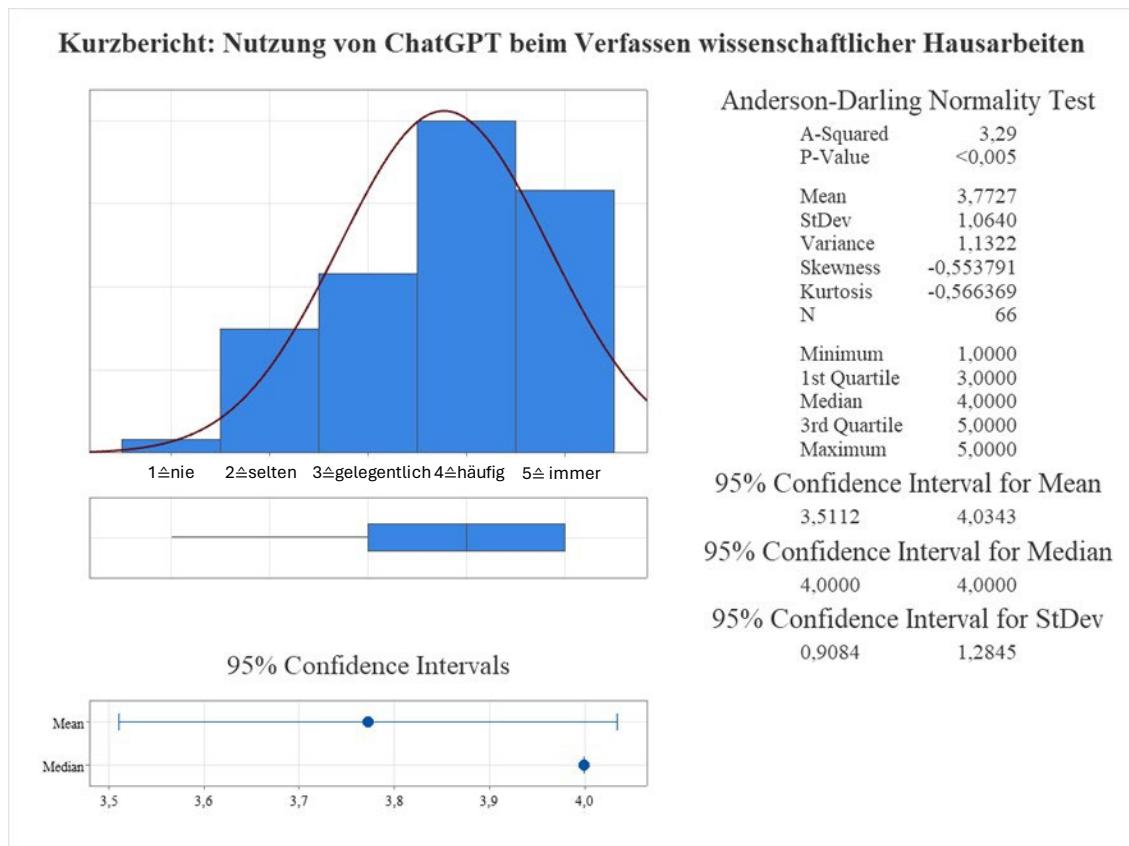


Abbildung 6.7: Kurzbericht: Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten

#### 6.4.2 Korrelationen

Die Pearson-Korrelationsmatrix ist eine statistische Methode, welche die Stärke und Richtung der linearen Beziehung zwischen zwei oder mehreren Variablen quantifiziert. Dabei berechnet sie den Pearson-Korrelationskoeffizienten ( $r$ ), der Werte zwischen -1 und 1 annehmen kann. Sofern  $r = 1$  ist, herrscht eine positive Korrelation, wenn die eine Variable steigt, steigt die andere ebenfalls. Wenn  $r = -1$  ist, dann sinkt eine Variable, sofern die andere steigt. Ist  $r = 0$ , so existiert keine Korrelation unter den Variablen (Field, 2009, S. 170ff.).

In Tabelle 1 ist eine Pearson-Korrelationsmatrix dargestellt mit fünf Variablen. Es ist zu erkennen, dass es zwei stark positive Korrelationen mit der Nutzung von ChatGPT gibt:

- Positive Korrelation zwischen der Erfahrung mit ChatGPT und der Nutzung von ChatGPT
- Positive Korrelation zwischen der Erfahrung mit ChatGPT und dem Einfluss auf die akademische Leistung in Form von Noten

Die Nutzung von ChatGPT korreliert niedriger positiv mit der Vertrauenswürdigkeit gegenüber den von ChatGPT generierten Informationen für akademische Arbeiten. Moderat positiv korreliert die Nutzung mit der Fehlerhäufigkeit der von ChatGPT bereitgestellten Informationen.

Die Vertrauenswürdigkeit, der von ChatGPT bereitgestellten Informationen hängt, moderat positiv mit der Wichtigkeit der Validität der Ergebnisse, welche von der Generativen KI generiert wurden, zusammen. Die Korrelation der Vertrauenswürdigkeit zum Einfluss auf die Leistung ist mit 0,1637 nur schwach positiv.

Die Häufigkeit der von ChatGPT generierten Fehler korreliert sehr niedrig positiv mit 0,1225 dem Einfluss auf die akademische Leistung der Studierenden.

	Erfahrung mit ChatGPT	Nutzung von ChatGPT	Vertrauens würdigkeit	Validität	Häufigkeit von Fehlern	Einfluss auf die Leistung
Erfahrung mit ChatGPT	1					
Nutzung von ChatGPT	0,491661	1				Dieser Bereich der Matrix ist leer, da sie genau den Kombinationen der unteren linken Diagonale entspricht.
Vertrauens würdigkeit	0,237065	0,150606	1			
Validität	0,116498	0,016489	0,340424	1		
Häufigkeit von Fehlern	0,193625	0,236300	-0,140766	0,216640	1	
Einfluss auf die Leistung	0,534390	0,350413	0,163695	0,267457	0,122472	1

Tabelle 6.1: Pearson Korrelationsmatrix

#### 6.4.3 Hypothesentests

Der Kruskal-Wallis-Test wird verwendet, um zu überprüfen, ob es signifikante Unterschiede zwischen den Medianen von drei oder mehr unabhängigen Gruppen gibt. Der Test ist eine nichtparametrische Methode, folglich trifft er keine Annahmen darüber, wie die Daten verteilt sind (McClenaghan, 2024). Da die vorliegenden Daten der Umfrage nicht normalverteilt sind, eignet sich der Kruskal-Wallis-Test, um die in Kapitel 2.9 aufgestellten Hypothesen zu bestätigen oder zu widerlegen.

Die Nullhypothese ( $H_0$ ), indiziert, dass die Medianen der Gruppen gleich sind. Die Alternativhypothese ( $H_1$ ) hingegen, indiziert, dass mindestens eine Gruppe einen anderen Median hat. Zudem ist auf den P-Wert zu achten, er gibt an, ob  $H_0$  oder  $H_1$  angenommen wird. Ein niedriger P-Wert, unter 0,05 weist darauf hin, dass  $H_0$  abgelehnt werden kann, was bedeutet, dass es signifikante Unterschiede zwischen den Gruppen gibt (McClenaghan, 2024).

Hypothese 1 und 2 aus Kapitel 2.9 beschreiben beide einen positiven Zusammenhang zwischen der Erfahrung mit ChatGPT und dem Vertrauen in Generative KI bei Studierenden (Amoozadeh, et al., 2024, S. 6; Acosta-Enriquez et al., 2024, S. 13ff.).

Tabelle 2 zeigt die Anzahl der Befragten Studierenden (N), den Median und die durchschnittliche Rangordnung für die Selbsteinschätzung der Erfahrung (von 1 bis 5). Der Z-Wert gibt an, wie viele Standardabweichungen ein Datenpunkt von dem Mittelwert einer Verteilung entfernt ist. Die Medianwerte variieren zwischen 2,0 und 3,0. Die Gruppen mit der höchsten Selbsteinschätzung

(Gruppe 5, 6 und 7) haben einen Median von 3,0, während die Gruppen mit der niedrigsten Selbsteinschätzung (Gruppe 1 und 4) einen Median von 2,0 aufweisen.

Erfahrung mit ChatGPT	N	Median	$\bar{\theta}$ Rang	Z-Wert
1	1	2,0	14,5	-1,00
2	2	2,5	25,3	-0,62
3	1	3,0	36,0	0,13
4	7	2,0	24,8	-1,27
5	20	3,0	34,0	0,15
6	21	3,0	32,1	-0,41
7	14	3,0	41,6	1,78
Gesamt	66		33,5	

Tabelle 6.2: Ergebnisse des Kruskal-Wallis-Tests zu den Hypothesen 1 und 2

In Tabelle 3 beträgt der H-Wert 5,44 ohne Berücksichtigung der Ränge, die gleich sind und 6,98 mit Berücksichtigung. Der Freiheitsgrade (DF) beträgt 6. Der P-Wert gibt an, wie wahrscheinlich es ist, dass die erhobenen Daten unter der Nullhypothese auftreten. Ein P-Wert von weniger als 0,05 wird als signifikant angesehen. Im vorliegenden Fall sind beide P-Werte deutlich größer als 0,05. Dies bedeutet, dass  $H_0$  angenommen wird. Folglich gibt es keinen signifikanten Unterschied in der Vertrauenswürdigkeit zwischen den Gruppen der Selbsteinschätzung der Erfahrung.

Nullhypothese ( $H_0$ )	$H_0$ : Alle Mediane sind gleich		
Alternativhypothese ( $H_1$ )	$H_1$ : Mindestens ein Median ist unterschiedlich		
Methode	DF	H-Value	P-Value
Nicht angepasst für Ties	6	5,44	0,489
Angepasst für Ties	6	5,98	0,426

Tabelle 6.3: Teststatistiken des Kruskal-Wallis Tests zu den Hypothesen 1 und 2

Hypothese 3 thematisiert die Bedenken hinsichtlich Plagiate und der Qualität der von ChatGPT generierten Inhalte. Sie behauptet, dass diese Variable einen negativen Einfluss auf das Vertrauen in diese Technologie haben (Obenza et al., 2023, S. 11).

Tabelle 4 zeigt, dass die Medianwerte für Gruppe 0 (keine Plagiatsgefahr) und Gruppe 2 (Plagiatsgefahr) gleich sind. Die  $\bar{\theta}$  Ränge sind ebenfalls sehr nah beieinander, was die Gleichheit der Mediane weiter unterstützt.

Plagiatsgefahr	N	Median	Ø Rang	Z-Wert
0	16	3	32,8	-0,16
2	50	3	33,7	0,16
Gesamt	66		33,5	

Tabelle 6.4: Ergebnis des Kruskal-Wallis-Tests zur Hypothese 3

Beide P-Werte in Tabelle 5 sind deutlich größer als 0,05. Dies bedeutet, dass H0 nicht abgelehnt werden kann. Es gibt keinen signifikanten Unterschied in der Vertrauenswürdigkeit in die von KI generierten Ergebnisse zwischen den Gruppen. Hypothese 3 ist damit widerlegt.

Nullhypothese (H0)	$H_0$ : Alle Mediane sind gleich		
Alternativhypothese (H1)	$H_1$ : Mindestens ein Median ist unterschiedlich		
Methode	DF	H-Value	P-Value
Nicht angepasst für Ties	1	0,02	0,875
Anangepasst für Ties	1	0,03	0,869

Tabelle 6.5: Teststatistiken des Kruskal-Wallis Tests zur Hypothese 3

Hypothese 4 behauptet, dass Studierende, die Generative KI kritisch analysieren und hinterfragen, ein höheres Maß an Vertrauen in die Technologie zeigen (Hu & Chan, 2023, S. 14f.). Diese Hypothese wird überprüft mittels der beiden Variablen, Validität und Vertrauenswürdigkeit.

Die Medianwerte in Tabelle 6 variieren zwischen 1,0 und 3,5. Die Gruppe mit der niedrigsten Erfahrung hat den niedrigsten Median, während die Gruppe mit einer moderaten Erfahrung einen Median von 3,5 hat. Der Z-Wert für Gruppe 1 beträgt -2,49, dies deutet darauf hin, dass diese Gruppe signifikant niedriger eingestuft wird als die anderen Gruppen.

Erfahrung mit ChatGPT	N	Median	$\bar{\theta}$ Rang	Z-Wert
1	3	1,0	6,5	-2,49
2	3	2,0	14,5	-1,75
3	6	3,5	42,7	1,23
4	25	3,0	32,3	-0,39
5	29	3,0	37,4	1,45
Gesamt	66		33,5	

Tabelle 6.6: Ergebnis des Kruskal-Wallis-Tests zur Hypothese 4

Beide P-Werte in Tabelle 7 sind kleiner als das übliche Signifikanzniveau von 0,05. Das bedeutet, dass H<sub>0</sub> abzulehnen ist. Es gibt einen signifikanten Unterschied in der Vertrauenswürdigkeit in die von ChatGPT generierten Ergebnisse zwischen den Gruppen mit unterschiedlicher Erfahrung. Dies unterstützt die Hypothese, dass Studierende, die Generative KI kritisch analysieren und hinterfragen, ein höheres Maß an Vertrauen in die Technologie zeigen.

Nullhypothese (H <sub>0</sub> )	H <sub>0</sub> : Alle Mediane sind gleich		
Alternativhypothese (H <sub>1</sub> )	H <sub>1</sub> : Mindestens ein Median ist unterschiedlich		
Methode	DF	H-Value	P-Value
Nicht angepasst für Ties	4	11,52	0,021
Angepasst für Ties	4	12,67	0,013

Tabelle 6.7: Teststatistiken des Kruskal-Wallis Tests zur Hypothese 4

#### 6.4.4 Multiple Regressionsanalyse

Die Regressionsgleichung zeigt, wie die Vertrauenswürdigkeit, in die von ChatGPT generierten Ergebnisse durch verschiedene Faktoren beeinflusst wird. Die allgemeine Form der Gleichung lautet:

Vertrauenswürdigkeit = Konstanten + Koeffizienten × Variablen

Nachfolgende Variablen gehen in die Gleichung ein:

Variablenkürzel	Variable
x <sub>1</sub>	Eigene Erfahrungen mit ChatGPT der Studierenden
x <sub>2</sub>	Häufigkeit der Nutzung von ChatGPT
x <sub>3</sub>	Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten
x <sub>4</sub>	Wichtigkeit von Validität der Ergebnisse von ChatGPT
x <sub>5</sub>	Beeinflussung durch Kommilitonen
x <sub>6</sub>	Fehlerhäufigkeit der von ChatGPT bereitgestellten Informationen
x <sub>7</sub>	Beeinflussung der akademischen Leistung durch ChatGPT

Tabelle 6.8: Variablenkürzel der Regressionsgleichungen

Die Koeffizienten in der Tabelle 9 zeigen den Einfluss jeder unabhängigen Variable auf die abhängige Variable. Ein positiver Koeffizient sorgt für einen Anstieg der unabhängigen Variable, welcher ebenso mit einem Anstieg der Vertrauenswürdigkeit einhergeht. Ein negativer Koeffizient indiziert, dass ein Anstieg der unabhängigen Variable mit einem Rückgang der Vertrauenswürdigkeit verbunden ist.

Geschlecht	Abhängige Variable	=	Konstante + Koeffizienten × Variablen
Männliches Geschlecht	Vertrauenswürdigkeit	=	1,844 + 0,040x <sub>1</sub> + 0,255x <sub>2</sub> + 0,210x <sub>3</sub> + 0,2914x <sub>4</sub> - 0,090x <sub>5</sub> - 0,2090x <sub>6</sub> - 0,228x <sub>7</sub>
Weibliches Geschlecht	Vertrauenswürdigkeit	=	1,052 + 0,040x <sub>1</sub> + 0,255x <sub>2</sub> + 0,210x <sub>3</sub> + 0,2914x <sub>4</sub> - 0,090x <sub>5</sub> - 0,2090x <sub>6</sub> - 0,228x <sub>7</sub>

Tabelle 6.9: Regressionsgleichungen

Ein p-Wert eines Koeffizienten, der kleiner als 0,05 ist, deutet darauf hin, dass der Faktor signifikant zur Erklärung der Variabilität des Vertrauens beiträgt. In den Ergebnissen der Tabelle 10 ist dies der Fall, für die „Wichtigkeit von Validität der Ergebnisse von ChatGPT“ und für die „Fehlerhäufigkeit der von ChatGPT bereitgestellten Informationen“. Dies bedeutet, dass diese Variable einen signifikanten Einfluss auf die Vertrauenswürdigkeit haben. Der Standardfehler des Koeffizienten gibt an, wie genau der geschätzte Koeffizient ist. Ein kleinerer Standardfehler deutet auf eine präzisere Schätzung hin. Folglich ist hier die Schätzung auch wieder präziser für die beiden zuvor hervorgehobenen Variablen.

Der t-Wert wird verwendet, um zu testen, ob der Koeffizient signifikant von null verschieden ist. Ein höherer t-Wert deutet darauf hin, dass die unabhängige Variable einen signifikanten Einfluss auf die abhängige Variable hat. Der VIF misst, wie stark die Varianz eines geschätzten Regressionskoeffizienten durch Multikollinearität, also der Korrelation zwischen den unabhängigen Variablen, erhöht wird. Ein VIF-Wert über 10 kann auf ein Problem mit Multikollinearität hinweisen. In der vorliegenden Tabelle 10 sind die VIF-Werte jedoch alle unter 3, weshalb kein ernsthaftes Problem in Bezug auf eine Multikollinearität besteht.

Bezeichnung	Koeffizient	Standardfehler des Koeffizienten	T-Wert	P-Wert	VIF
Konstante (Vertrauenswürdigkeit)	1,844	0,770	2,40	0,020	
Eigene Erfahrungen mit ChatGPT der Studierenden	0,040	0,109	0,37	0,712	2,37
Häufigkeit der Nutzung von ChatGPT	0,255	0,183	1,39	0,169	2,24
Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten	0,210	0,117	1,80	0,077	1,86
Wichtigkeit von Validität der Ergebnisse von ChatGPT	0,2914	0,0978	2,98	0,004	1,29
Beeinflussung durch Kommilitonen	-0,090	0,111	-0,81	0,422	1,36
Fehlerhäufigkeit der von ChatGPT bereitgestellten Informationen	-0,2090	0,0809	-2,58	0,012	1,25
Beeinflussung der akademischen Leistung durch ChatGPT	-0,228	0,143	-1,59	0,116	1,77
Geschlecht					
2	-0,792	0,231	-3,43	0,001	1,38
3	-1,13	1,18	-0,96	0,342	2,54

Tabelle 6.10: Koeffizienten der Regressionsgleichungen

Tabelle 11 bewertet die Qualität der multiplen Regressionsanalyse, dabei gibt „S“ die Standardabweichung der Residuen an. Dieser Wert liegt beim vorliegenden Modell bei 0,73 und gibt die durchschnittliche Abweichung der beobachteten Werte von den vorhergesagten Werten an. Der eher niedrige Wert deutet somit auf eine gute Anpassung des Modells hin.

Das Bestimmtheitsmaß („R-sq“) gibt an, wie viel der Varianz in der abhängigen Variablen durch die unabhängigen Variablen erklärt wird. Ein R-Quadrat von 47,02% bedeutet, dass das Modell etwa 47 % der Varianz der abhängigen Variablen (Vertrauenswürdigkeit) erklären kann, was als mäßig angesehen werden kann. Folglich gibt es noch andere Faktoren, welche die Vertrauenswürdigkeit beeinflussen.

Der angepasste R-Quadrat-Wert („R-sq(adj)“) von 38,50 % berücksichtigt die Anzahl der Variablen im Modell und ist nützlich, um die Modellgüte zu bewerten.

S	R-sq	R-sq(adj)
0,734570	47,02%	38,50%

Tabelle 6.11: Model Zusammenfassung der Regression

Tabelle 12 zeigt die Ergebnisse der Analyse der Varianz für die multiple Regressionsanalyse. Die angepasste Summe der Quadrate (Adj SS) erklärt die Variation, welche durch die jeweiligen Quellen verursacht wird. Höhere Werte deuten dabei auf eine größere Erklärung der Variation hin. „Adj MS“

gibt die angepasste Summe der Quadrate geteilt durch die Freiheitsgrade an und zeigt die mittlere Variation, die durch jede Quelle erklärt wird.

Der F-Wert ist das Verhältnis der mittleren Quadrate der Regression zu den mittleren Quadraten des Fehlers. Der F-Wert von 5,52 zeigt, dass das Modell insgesamt signifikant ist. Ein höherer F-Wert deutet darauf hin, dass die unabhängigen Variablen signifikant zur Erklärung der Variation in der abhängigen Variablen beitragen.

Die gesamte Regression hat einen p-Wert von 0, was bedeutet, dass das Modell insgesamt signifikant ist und die unabhängigen Variablen die abhängige Variable signifikant erklären. Die Variable „Eigene Erfahrungen mit ChatGPT der Studierenden“ hat einen p-Wert von 0,712, was darauf hinweist, dass sie nicht signifikant zur Erklärung der Variation beiträgt. Die Häufigkeit der Nutzung von ChatGPT, die Beeinflussung durch Kommilitonen und die Beeinflussung der akademischen Leistung durch ChatGPT sind ebenfalls nicht signifikant. Die Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten ist mit einem p-Wert von 0,077 grenzwertig signifikant, könnte aber in einigen Kontexten als signifikant betrachtet werden. Die Wichtigkeit von Validität der Ergebnisse von ChatGPT und die Fehlerhäufigkeit der von ChatGPT bereitgestellten Informationen tragen signifikant zur Erklärung der Variation bei. Der p-Wert des Geschlechts von 0,005 deutet darauf hin, dass diese Variable ebenfalls einen signifikanten Einfluss auf die abhängige Variable hat.

Der p-Wert für den Fehler zeigt an, dass 30,2172 der Variation nicht durch die unabhängigen Variablen erklärt wird. Der Lack-of-Fit und der Reiner Fehler zeigt an, dass es keinen signifikanten Mangel an Anpassung gibt.

Quelle	DF	Adj SS	Adj MS	F-Wert	P-Wert
Regression	9	26,8131	2,97923	5,52	0,000
Eigene Erfahrungen mit ChatGPT der Studierenden	1	0,0744	0,07444	0,14	0,712
Häufigkeit der Nutzung von ChatGPT	1	1,0482	1,04818	1,94	0,169
Nutzung von ChatGPT beim Verfassen wissenschaftlicher Hausarbeiten	1	1,7468	1,74676	3,24	0,077
Wichtigkeit von Validität der Ergebnisse von ChatGPT	1	4,7936	4,79358	8,88	0,004
Beeinflussung durch Kommilitonen	1	0,3529	0,35287	0,65	0,422
Fehlerhäufigkeit der von ChatGPT bereitgestellten Informationen	1	3,6057	3,60568	6,68	0,012
Beeinflussung der akademischen Leistung durch ChatGPT	1	1,3720	1,37202	2,54	0,116
Geschlecht	2	6,3912	3,19560	5,92	0,005
Error	56	30,2172	0,53959		
Lack-of-Fit	55	30,2172	0,54940	*	*
Pure Error	1	0,0000	0,00000		
Gesamt	65	57,0303			

Tabelle 6.12: Analyse der Varianz der Regressionsgleichungen

## 6.5 Diskussion

Die Pearson-Korrelationsmatrix aus Tabelle 1 zeigt, dass die Erfahrung mit ChatGPT und die Nutzung des Tools signifikante positive Zusammenhänge mit dem Einfluss auf die Leistung der Studierenden aufweisen. Dies deutet darauf hin, dass Studierende, die mehr Erfahrung mit ChatGPT haben, auch eher bereit sind, es häufiger zu nutzen. Und dass erfahrene Nutzer die Ergebnisse von ChatGPT effektiver in ihren akademischen Arbeiten verwenden.

Die niedrige Korrelation zwischen Nutzung und Vertrauen zeigt, dass viele Studierende ChatGPT verwenden, aber an der Qualität zweifeln. Die Vertrauenswürdigkeit der Ergebnisse haben eine moderate Korrelation mit der Validität, was darauf hindeutet, dass das Vertrauen in die Technologie möglicherweise durch die Wahrnehmung ihrer Validität beeinflusst wird. Wenn Studierende folglich glauben, dass die von ChatGPT generierten Ergebnisse valide sind, könnten sie eher dazu geneigt sein, den Ergebnissen zu vertrauen. Umgekehrt könnte ein Mangel an Vertrauen in die Validität der Ergebnisse zu einem geringeren Vertrauen in die Technologie führen. Die wahrgenommenen Fehler scheinen jedoch nur einen geringen Einfluss auf die akademische Leistung zu haben, was darauf hinweist, dass Studierende trotz der Wahrnehmung von Fehlern in den Ergebnissen möglicherweise weiterhin auf ChatGPT vertrauen und es in ihrem Studium verwenden. Das bedeutet, dass Studierende, die die Ergebnisse von ChatGPT als vertrauenswürdig empfinden, auch eher die Validität dieser Ergebnisse anerkennen.

Die Hypothesen 1 und 2 aus Kapitel 2.9 können mithilfe der Auswertung aus Kapitel 4.2 bestätigt werden. Studierende, die ihre Erfahrung mit ChatGPT höher einschätzen, haben tendenziell eine höhere Vertrauenswürdigkeit, in die von ChatGPT generierten Ergebnisse. Somit führt eine höhere Selbsteinschätzung zu einem höheren Vertrauen in die Technologie. Diese Schlussfolgerung unterstützt auch die Korrelationsmatrix aus Tabelle 1.

Die Ergebnisse des Kruskal-Wallis-Tests aus Tabelle 4 der Hypothese 3 aus Kapitel 2.9 widerlegen jedoch die These. Es gibt keinen signifikanten Zusammenhang zwischen der Plagiatsgefahr und der Vertrauenswürdigkeit in die von ChatGPT generierten Inhalte. Die Studierenden scheinen unabhängig von ihren Bedenken hinsichtlich Plagiate ein ähnliches Vertrauen in die Technologie zu haben.

Hypothese 4 aus Kapitel 2.9 kann hingegen mithilfe der Auswertung aus Kapitel 4.2 bestätigt werden. Die Erfahrung mit ChatGPT hat einen signifikanten Einfluss auf die Wahrnehmung der Vertrauenswürdigkeit. Je höher die Erfahrung, desto höher das Vertrauen in die Technologie. Diese Bestätigung der Hypothese 4 unterstützt auch die Korrelationsmatrix aus Tabelle 1.

Die Ergebnisse der multiplen Regressionsanalyse unterstützen die Hypothese, dass verschiedene Faktoren, wie die Selbsteinschätzung der Erfahrung und die Nutzung von ChatGPT, signifikant zum Vertrauen in Generative KI beitragen. Insbesondere zeigt die Variable „Eigene Erfahrungen mit ChatGPT der Studierenden“ einen signifikanten Einfluss auf die Vertrauenswürdigkeit. Die Analyse legt nahe, dass das Vertrauen in Generative KI durch die kritische Auseinandersetzung mit der Technologie und die persönliche Erfahrung der Studierenden gestärkt werden kann. Die Ergebnisse zeigen, dass das Modell eine moderate Fähigkeit hat, die Variation der abhängigen Variable zu erklären, da das R-Quadrat bei 47,02% liegt. Dies bedeutet, dass fast die Hälfte der Variabilität durch die unabhängigen Variablen erklärt wird, was darauf hinweist, dass es möglicherweise noch andere Faktoren gibt, die nicht im Modell enthalten sind und die ebenfalls die abhängige Variable beeinflussen könnten. Die Standardabweichung der Residuen von 0,734570 ist niedrig, was darauf hindeutet, dass die Vorhersagen des Modells im Allgemeinen nah an den tatsächlichen Werten liegen. Das bereinigte R-Quadrat von 38,50% zeigt, dass das Modell zwar einige Erklärungen bietet,

aber auch, dass die Hinzunahme zusätzlicher Variablen möglicherweise nicht zu einer signifikanten Verbesserung der Modellanpassung führt.

Insgesamt weist Tabelle 12 darauf hin, dass das Modell nützlich ist, aber Raum für Verbesserungen und weitere Analysen besteht, um die Vorhersagekraft zu erhöhen. Zudem zeigt die Analyse, dass einige der unabhängigen Variablen signifikant zur Erklärung der Variation der abhängigen Variablen beitragen, insbesondere die Wichtigkeit der Validität der Ergebnisse von ChatGPT, die Fehlerhäufigkeit und das Geschlecht. Andere Variablen, wie die eigenen Erfahrungen der Studierenden und die Häufigkeit der Nutzung, sind nicht signifikant. Das Modell als Ganzes ist signifikant, was darauf hindeutet, dass es nützlich ist, um die abhängige Variable zu erklären.

## 6.6 Fazit

Zusammenfassend lässt sich festhalten, dass das Vertrauen der Studierenden in ChatGPT stark von der persönlichen Erfahrung und der kritischen Auseinandersetzung mit der Technologie abhängt. Die Ergebnisse der quantitativen Analyse zeigen, dass Studierende, die über umfangreiche Erfahrungen mit ChatGPT verfügen, tendenziell ein höheres Vertrauen in die Validität der generierten Inhalte aufweisen. Dies unterstreicht die Bedeutung einer fundierten Ausbildung im Umgang mit digitalen Technologien.

Die Erkenntnisse dieser Arbeit haben weitreichende Implikationen für die Praxis. Hochschulen sollten gezielte Schulungsangebote entwickeln, um Studierenden die notwendigen Kompetenzen im Umgang mit generativer KI zu vermitteln. Eine kritische Reflexion der von KI generierten Inhalte sollte Bestandteil der Lehrpläne werden, um das Vertrauen in diese Technologien nachhaltig zu stärken.

Zukünftige Forschungen könnten sich darauf konzentrieren, die langfristigen Auswirkungen der Nutzung von generativer KI auf die akademische Leistung der Studierenden zu untersuchen. Darüber hinaus wäre es sinnvoll, den Einfluss von sozialen Faktoren, wie dem Austausch von Erfahrungen unter Kommilitonen, auf das Vertrauen in KI-gestützte Technologien näher zu beleuchten. Solche Studien könnten wertvolle Erkenntnisse für die Weiterentwicklung von Lehrmethoden und -inhalten im digitalen Zeitalter liefern.

## 6.7 Literaturverzeichnis

- Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., Mohammed, A. H., Hassan, B. A. R., Wayyes, A. M., Farhan, S. S., Khatib, S. E., Rahal, M., Sahban, A., Abdelaziz, D. H., Mansour, N. O., AlZayer, R., Khalil, R., Fekih-Romdhane, F., Hallit, R., ... Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-52549-8>
- Acosta-Enriquez, B. G., Ballesteros, M. A. A., Jordan, O. H., Roca, C. L. & Tirado, K. S. (2024). Analysis of college students' attitudes toward the use of ChatGPT in their academic activities: effect of intent to use, verification of information and responsible use. *BMC Psychology*, 12(1). <https://doi.org/10.1186/s40359-024-01764-z>
- Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Ragavan, S. S. & Alipour, M. A. (2024). Trust in Generative AI among Students: An exploratory study. *BMC Psychology*, 12(1), 255. <https://doi.org/10.1145/3626252.3630842>
- Bortz, J. & Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler. In Springer-Lehrbuch. <https://doi.org/10.1007/978-3-642-12770-0>
- Field, A. (2009). Discovering statistics using SPSS. SAGE Publications.

- Flanagin, A. J. & Metzger, M. J. (2008). Digital media and youth: Unparalleled opportunity and unprecedented responsibility. 8 Massachusetts Institute Of Technology.  
<http://mobilelearningcourse.pbworks.com/f/Digital%20Media%20and%20Youth.pdf>
- Hu, K. (2023, 2. Februar). ChatGPT sets record for fastest-growing user base—Analyst note. Reuters. Abgerufen am 27. April 2025, von <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
- Lund, B. D. & Wang, T. (2023). Chatting about ChatGPT: how may AI and GPT impact academia and libraries? Library Hi Tech News, 40(3), 26–29. <https://doi.org/10.1108/lhtn-01-2023-0009>
- McClenaghan, E. (2024, 3. Mai). The Kruskal-Wallis Test. Informatics From Technology Networks. <https://www.technologynetworks.com/informatics/articles/the-kruskal-wallis-test-370025>
- McKnight, D. H., Choudhury, V. & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. Information Systems Research, 13(3), 334–359. <https://doi.org/10.1287/isre.13.3.334.81>
- Menon, D. & Shilpa, K. (2023). “Chatting with ChatGPT”: Analyzing the factors influencing users’ intention to Use the Open AI’s ChatGPT using the UTAUT model. Heliyon, 9(11), e20962. <https://doi.org/10.1016/j.heliyon.2023.e20962>
- Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. Journal Of The American Society For Information Science And Technology, 58(13), 2078–2091. <https://doi.org/10.1002/asi.20672>
- Mhlanga, D. (2023). Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4354422>
- Multiple regressionsanalyse. (o. D.). Methodenberatung | UZH. [https://www.methodenberatung.uzh.ch/de/datenanalyse\\_spss/zusammenhaenge/mreg.html](https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/zusammenhaenge/mreg.html)
- Nemt-Allah, M., Khalifa, W., Badawy, M., Elbably, Y. & Ibrahim, A. (2024). Validating the ChatGPT Usage Scale: psychometric properties and factor structures among postgraduate students. BMC Psychology, 12(1). <https://doi.org/10.1186/s40359-024-01983-4>
- Obenza, B. N., Salvahan, A., Rios, A. N., Solo, A., Alburo, R. A. & Gabila, R. J. (2023). University Students’ Perception and Use of ChatGPT: Generative Artificial Intelligence (AI) in Higher Education. INTERNATIONAL JOURNAL OF HUMAN COMPUTING STUDIES, 5(12). <https://journals.researchparks.org/index.php/IJHCS>
- OpenAI. (2022). ChatGPT: Optimizing Language Models for Dialogue. OpenAI Blog. Abgerufen am 10. Januar 2023, von <https://openai.com/blog/chatgpt/>
- OpenAI. (2023, 14. März). GPT-4. Abgerufen am 17. August 2023, von <https://openai.com/index/gpt-4-research/>
- Schuster, T. & Liesen, A. (2017). Statistik für Wirtschaftswissenschaftler. In Springer eBooks. <https://doi.org/10.1007/978-3-662-49836-1>
- Slade, E. L., Dwivedi, Y. K., Piercy, N. C. & Williams, M. D. (2015). Modeling Consumers’ Adoption Intentions of Remote Mobile Payments in the United Kingdom: Extending UTAUT with Innovativeness, Risk, and Trust. Psychology And Marketing, 32(8), 860–873. <https://doi.org/10.1002/mar.20823>

## 7 KI in der Prüfungsvorbereitung

### *Die Auswirkungen der Nutzung generativer KI auf die Prüfungsangst bei Studierenden im Kontext der Klausurvorbereitung*

Angela Altomare

Paula Vanselow

#### **Abstract**

In der vorliegenden Studie wird untersucht, wie die Nutzung von generativer KI und insbesondere ChatGPT sich auf das Erleben von Prüfungsangst von Studierenden auswirkt. Durch die zunehmende Verbreitung von generativer KI im Hochschulkontext wird der Frage nachgegangen, ob der Einsatz solcher Technologien die Prüfungsangst reduzieren kann.

Zur Untersuchung der Zusammenhänge zwischen der Häufigkeit der Nutzung von ChatGPT und verschiedenen Dimensionen der Prüfungsangst (affektiv, kognitiv, physiologisch) wurden Daten von Studierenden des Bachelorstudiengangs Wirtschaftsinformatik der Frankfurt University of Applied Sciences analysiert.

Die Befunde deuten auf schwache, nicht signifikante Zusammenhänge zwischen der Nutzung von ChatGPT und Prüfungsangst hin. Jedoch konnten geschlechterspezifische Unterschiede in der affektiven und physiologischen Komponente sowie signifikante Differenzen in der Bewertung der Effektivität von ChatGPT in Abhängigkeit von Studiensemestern festgestellt werden. Diese Ergebnisse lassen darauf schließen, dass der Einsatz generativer KI Studierende Unterstützung bieten kann, jedoch kann kein direkter Zusammenhang zu einer Reduzierung der Prüfungsangst nachgewiesen werden. Zuletzt gibt die Studie einen Ausblick auf die notwendigen Kompetenzen im Umgang mit KI und mögliche Implikationen für die Hochschullehre.

**Schlagwörter:** ChatGPT, Prüfungsangst, generative Künstliche Intelligenz, Hochschulbildung, Studierendenverhalten, Klausurvorbereitung, Kompetenz

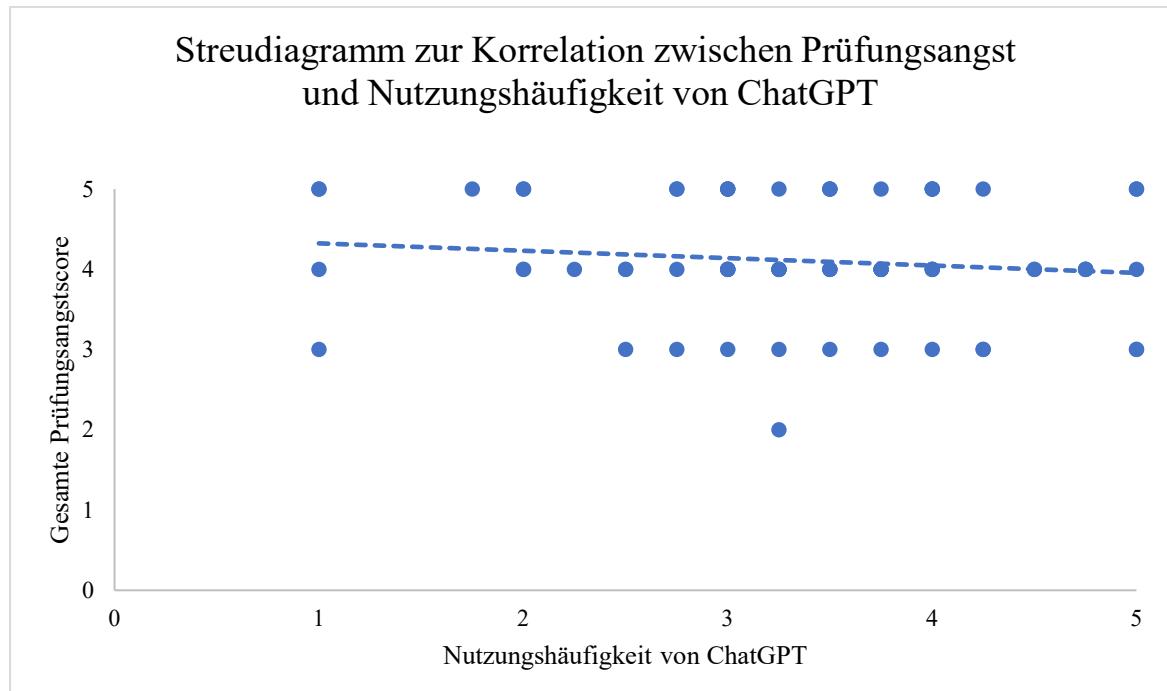


Abbildung 7.1 Zusammenhang zwischen Prüfungsangst und Nutzungshäufigkeit von ChatGPT

## 7.1 Themenhintergrund und Zielsetzung

In diesem Kapitel wird zunächst die Relevanz des Themas der Forschungsarbeit erläutert. Anschließend werden die Zielsetzung und Problemstellung dargestellt, die sich aus dem Einsatz generativer KI in der Hochschulbildung und der Prüfungsangst ergibt. Abschließend wird der Aufbau der Arbeit dargestellt, um einen Überblick über den inhaltlichen Verlauf der Arbeit zu geben.

### 7.1.1 Themenrelevanz

Der zunehmende Einsatz generativer KI verändert die Hochschulbildung grundlegend. Insbesondere Tools wie ChatGPT werden von Studierende immer häufiger zur Klausurvorbereitung genutzt (von Garrel & Mayer, 2023, S.7). Diese Entwicklung betrifft nicht nur den Lern- und Lehrprozess, sondern auch die psychologischen Aspekte des Studiums.

Ein bislang kaum untersuchter Zusammenhang besteht zwischen der Nutzung von KI-Tools und emotionalen Faktoren wie Prüfungsangst. Dabei zählt Prüfungsangst zu den häufigsten Herausforderungen im Studium und kann Motivation, Leistungsfähigkeit und das Wohlbefinden erheblich beeinflussen (Zeidner, 1998, S.4). In diesem Zusammenhang stellt sich die Frage, ob generative KI Studierende in Prüfungssituationen entlasten kann.

Gleichzeitig stellt der Umgang mit KI-Tools neue Anforderungen an die Studierenden. Künftig gewinnen sowohl technische Kompetenzen als auch das kritische Denken und die Fähigkeit, KI-generierte Inhalte einzuordnen und zu bewerten, zunehmend an Bedeutung (Luckin et al., 2016, S. 11). Somit hängt der mögliche Nutzen von den individuellen Kompetenzen der Studierenden ab.

Die vorliegende Studie greift diese Entwicklung auf und beleuchtet den möglichen Zusammenhang zwischen ChatGPT Nutzung und Prüfungsangst im Hochschulkontext.

### **7.1.2 Zielsetzung der Arbeit und Problemstellung**

Prüfungsangst ist ein weit verbreitetes Phänomen im Hochschulkontext und kann sich negativ auf das psychische Wohlbefinden sowie die Lern- und Leistungsfähigkeit von Studierenden auswirken. (Pekrun & Goetz, 2006, S.248).

Gleichzeitig verändern generative KI-Tools wie ChatGPT die Lerngewohnheiten von Studierenden und wie sie sich auf Prüfungen vorbereiten (Sousa & Cardoso, 2025, S.1).

Bislang ist wenig darüber bekannt, ob generative KI lediglich ein weiteres Hilfsmittel für Studierende darstellt oder ob sie tatsächlich einen Einfluss auf Prüfungsangst ausüben kann.

Ziel der vorliegenden Arbeit ist es somit, Erkenntnisse über den Zusammenhang zwischen der Nutzung generativer KI, insbesondere ChatGPT, in der Klausurvorbereitung und dem Erleben von Prüfungsangst bei Studierenden zu gewinnen. Darüber hinaus wird auch untersucht, ob sich durch die Nutzung von ChatGPT die individuellen Prüfungsvorbereitung verändert und welche neuen Kompetenzanforderungen sich daraus möglicherweise ableiten lassen.

### **7.1.3 Forschungsfrage und Hypothesen**

Vor diesem Hintergrund stellt sich die zentrale Forschungsfrage dieser Arbeit wie folgt: Welche Auswirkung hat die Nutzung generativer KI-Tools wie ChatGPT auf die Prüfungsangst im Kontext der Klausurvorbereitung?

Auf Basis der formulierten Fragestellung werden folgende Hypothesen aufgestellt:

H1: Die Nutzung generativer KI steht in negativem Zusammenhang mit der Prüfungsangst von Studierenden

H2: Es bestehen geschlechterspezifische Unterschiede in der Ausprägung der Prüfungsangst bei der Klausurvorbereitung bei der Nutzung von ChatGPT

H3: Die wahrgenommene Prüfungsangst bei der Klausurvorbereitung mit ChatGPT unterscheidet sich je nach Studiensemester

Diese Hypothesen bilden die Grundlage für die empirische Untersuchung im Rahmen dieser Arbeit.

### **7.1.4 Aufbau des Artikels**

Die vorliegende Arbeit gliedert sich in mehrere zentralen Abschnitte, die im Folgenden näher erläutert werden.

Nach der Einleitung folgt der theoretische Rahmen, in dem relevanten thematischen Inhalte eingeführt werden und Zusammenhänge zu Prüfungsangst, generative KI und deren Einsatz im Hochschulkontext erläutert werden.

Im Anschluss erfolgt ein Überblick des aktuellen Forschungsstands sowie die Identifizierung von Forschungslücken. Nach einer ausführlichen Darstellung des methodischen Hintergrunds werden die Ergebnisse der Datenerhebung im weiteren Verlauf detailliert dargestellt. Anschließend werden die Ergebnisse im Hinblick auf die Forschungsfrage, die Praxis und weitere Forschungsarbeiten diskutiert.

Im Fazit werden die wesentlichen Ergebnisse noch einmal kurz dargestellt und es wird explizit auf die Beantwortung der Forschungsfrage und die Hypothesen eingegangen.

## **7.2 Grundlagen der Prüfungsangst und Veränderungen der Kompetenzen: Theoretische Modelle und Forschungslücken**

In diesem Kapitel werden die zentralen theoretischen Grundlagen erläutert. Dazu zählen Modelle zur Prüfungsangst sowie zentrale Einflussfaktoren. Abschließend erfolgt eine Darstellung des Forschungsstands sowie der Forschungslücken.

### **7.2.1 Prüfungsangst: Definitionen und Einflussfaktoren**

Prüfungsangst stellt einen emotionalen Zustand dar, der vorrangig im schulischen und akademischen Kontext auftritt und in der Forschung als eine besondere Form der Leistungsangst angesehen wird. Sie äußert sich primär in Prüfungssituationen und hat häufig einen negativen Einfluss auf die Leistungsfähigkeit (Spielberger & Vagg, 1995, S.3-5). Menschen mit Prüfungsangst lassen sich in Prüfungssituationen leichter ablenken, haben häufiger Schwierigkeiten, einfache Anweisungen zu verstehen und können relevante Informationen nicht abrufen (Zeidner, 1998, S. 4).

Studien zeigen, dass hohe Selbstwirksamkeit mit geringerer Prüfungsangst einhergeht, da mehr Vertrauen in die eigenen Fähigkeiten besteht (Maier et al., 2021, S. 2). Auch hohe familiäre oder eigene Erwartungen sowie starker Leistungsdruck können Prüfungsangst fördern (Zheng et al., 2023, S. 2). Externe Rahmenbedingungen verstärken sie zusätzlich, vor allem wenn Prüfungen stark gewichtet sind und wenig Kontrolle besteht (Zeidner, 1998, S. 4).

In der Literatur lässt sich keine allgemeingültige Definition von Prüfungsangst feststellen. Stattdessen existieren zahlreiche Theorien und Ansätze, die das Konstrukt aus unterschiedlichen Perspektiven beleuchten. Während manche Autoren Prüfungsangst als eine zustandsspezifische Reaktion definieren (Spielberger & Vagg, 1995, S. 3), verstehen andere das Konstrukt als mehrdimensional (Stöber, 2004, S. 213; Liebert & Morris, 1967, S. 975–976). Das Modell von Liebert und Morris (1967, S.975–976) unterscheidet zwischen einer kognitiven und emotionalen Komponente. Die kognitive Komponente (“worry”) umfasst Sorgen über die eigene Fähigkeit eine gute Leistung zu bringen. Die emotionale Komponente (“emotionality”) hingegen bezieht sich vor allem auf physiologische Reaktionen wie Zittern, Schwitzen oder Herzklopfen (Liebert & Morris, 1967, S.975–976). In der Weiterentwicklung des Modells wurde betont, dass die ursprüngliche „emotionality“-Komponente nicht nur physiologische Reaktionen, sondern auch subjektiv empfundene emotionale Zustände beinhaltet (Cassady & Johnson, 2002, S.273).

Das Modell von Liebert und Morris bildet die Grundlage für das Test Anxiety Inventory (TAI). Dieses Instrument ermöglicht, sowohl kognitive Sorgen als auch physiologische und affektive Reaktionen zu messen und bildet deswegen die theoretische Grundlage dieser Studie (Spielberger, 1980, S. 1f.).

Zusammenfassend lässt sich sagen, dass Prüfungsangst ein komplexes, mehrdimensionales Phänomen ist, das sowohl von individuellen als auch von situativen Faktoren beeinflusst wird und erhebliche Auswirkungen auf die Leistungsfähigkeit haben kann.

### **7.2.2 Veränderungen der Kompetenzen durch generative KI in der Hochschulbildung**

Obwohl unterschiedliche KI-Tools in Hochschulen verwendet werden, fokussiert sich diese Arbeit ausschließlich auf ChatGPT, aufgrund der hohen Verbreitung, einfachen Zugänglichkeit und den breiten Bekanntheitsgrad unter Studierenden.

Die Nutzung von generativer KI bringt veränderte Anforderungen an die Studierenden mit sich. Um generative KI-Tools sinnvoll nutzen zu können, sind nicht nur grundlegende technische Kenntnisse nötig, sondern auch kritisches Denken, da KI generierte Inhalte fehlerhafte Informationen oder auch Verzerrungen und diskriminierende Tendenzen enthalten können (Educational Technology Journal, 2024, S.3; Alasadi & Baiz, 2023, S.2967). Besonders herausfordernd ist dabei die Abgrenzung von Eigenleistung und KI-Unterstützung, da die Grenzen zunehmend verschwimmen (Schmohl et al., 2023, S.10 ff.).

Insgesamt wird somit deutlich, dass insbesondere ChatGPT eine zentrale Rolle im Hochschulbereich einnimmt. Es bietet weitreichende Möglichkeiten zur Unterstützung von Studierenden, verlangt jedoch auch einen gewissenhaften und reflektierten Umgang sowie die Entwicklung neuer Kompetenzen der Studierenden.

### 7.2.3 Forschungsstand und Forschungslücke

Im Folgenden werden einige der bisherigen Studien zum Thema generative KI im Hochschulbereich, insbesondere in Bezug auf Prüfungsangst und Klausurvorbereitung, betrachtet. In diesem Zusammenhang wird auch auf die Forschungslücken eingegangen, um die Relevanz der vorliegenden Arbeit hervorzuheben.

Aufgrund des begrenzten Umfangs der Arbeit werden hier lediglich einige ausgewählte, aktuelle Studien berücksichtigt.

In einer Studie aus dem Jahr 2025 wurde im Rahmen einer vierwöchigen Untersuchung der Einfluss der Nutzung von ChatGPT auf die Prüfungsangst und die akademische Leistung von Pharmaziestudierende analysiert. Eine Versuchsgruppe verwendete ChatGPT zum Lernen, während die Kontrollgruppe lediglich die üblichen Anweisungen erhielt. Die Auswertung der Ergebnisse ergab, dass sich weder die Prüfungsergebnisse noch das Ausmaß der Prüfungsangst signifikant zwischen den Gruppen unterschieden. Demnach kommt die Studie zu dem Schluss, dass KI-Tools kurzfristig keinen messbaren Einfluss auf Prüfungsleistung und Prüfungsangst haben. Gleichzeitig betonen die Autoren, dass in zukünftigen Studien langfristige Effekte untersucht werden sollten (Ali et al., 2025, S.2).

Eine weitere Studie aus dem Jahr 2024 untersuchte den Einfluss der Nutzung von ChatGPT auf die Prüfungsangst und auf emotionale Reaktionen von 160 Studierenden. Die Ergebnisse zeigen, dass der Einsatz von ChatGPT sowohl Prüfungsangst als auch negative Emotionen spürbar reduzieren kann, während positive Emotionen verstärkt werden. Damit wird deutlich, dass KI-Anwendungen wie ChatGPT eine emotionale Entlastung für Studierende bieten (Gao, 2024, S.199).

Obwohl erste Studien bereits Hinweise auf mögliche Auswirkungen der Verwendung generativer KI auf Einflussfaktoren wie Prüfungsangst liefern, ist die Forschung in diesem Bereich insgesamt noch begrenzt. Es fehlt die Erkenntnis darüber, inwiefern sich der Einsatz generativer KI im Rahmen der Klausurvorbereitung konkret auf das Erleben von Prüfungsangst von Studierenden auswirkt. Zudem gibt es bislang nur wenige Studien, die untersuchen, inwiefern sich die Klausurvorbereitung von Studierenden mit ChatGPT-Nutzung von derjenigen ohne den Einsatz solcher Tools unterscheidet und welche neuen Kompetenzen dafür erforderlich sind.

Die vorliegende Untersuchung hat zum Ziel, bestehende Forschungslücken zu schließen und fundierte Erkenntnisse über den Zusammenhang zwischen der Nutzung generativer KI in der Klausurvorbereitung und dem Erleben von Prüfungsangst bei Studierenden zu gewinnen. Darüber hinaus werden die daraus resultierenden Veränderungen in den Lernstrategien sowie die neuen erforderlichen Kompetenzen der Studierenden analysiert.

## **7.3 Auf dem Weg zur Erkenntnis**

Im folgenden Kapitel wird das methodische Vorgehen der Studie beschrieben, einschließlich der Datenerhebung und -auswertung. Weiterhin werden auch das Erhebungsinstrument und die Operationalisierung zentraler Konzepte erläutert. Abschließend werden die Gütekriterien zur Sicherung der wissenschaftlichen Qualität dargestellt. Die detaillierte Vorgehensweise der Datenerhebung und Stichprobe dieser Arbeit wird bereits im allgemeinen Kapitel zur Methodik unter 1.2 näher betrachtet und entfällt daher an dieser Stelle.

### **7.3.1 Forschungslücke**

Trotz der zunehmenden Relevanz generativer KI im Hochschulbereich existieren bislang nur wenige empirische Studien, die die konkrete Nutzung solcher Technologien in der Prüfungsvorbereitung und deren Auswirkungen auf Faktoren wie Prüfungsangst analysieren. Auch die Frage, welche neuen Kompetenzen im Umgang mit KI erforderlich sind, wurde bisher kaum untersucht. Die vorliegende Studie leistet daher einen Beitrag zur Schließung dieser Forschungslücke und liefert erste Erkenntnisse über studentische Erfahrungen mit KI-gestütztem Lernen.

### **7.3.2 Erhebungsinstrumente und Operationalisierung**

Die Datenerhebung erfolgte mittels eines selbsterstellten, standardisierten Online-Fragebogens. Der Fragebogen bestand aus insgesamt 34 geschlossenen Fragen bzw. Aussagen, die mithilfe der Likert-Skala bewertet wurden. Diese Rating-Skala ermöglicht eine differenzierte Einschätzung von Meinungen, Einstellungen und Verhalten, da die Teilnehmenden aus abgestuften Antwortkategorien wählen können (Steiner & Benesch, 2018, S.58).

Der erste Themenblock diente der Erfassung soziodemografischer Merkmale wie Alter, Geschlecht und aktuelles Studiensemester sowie die bisherige Nutzungserfahrung mit ChatGPT im universitären Kontext. Zusätzlich wurden die Teilnehmenden nach der Häufigkeit der Nutzung von ChatGPT im Rahmen der Klausurvorbereitung erfragt.

Ein zentraler Bestandteil des Fragebogens war die Messung von Prüfungsangst. Diese wurde mithilfe einer fünfstufigen Likert-Skala erfasst. Die zugrundeliegenden Aussagen orientieren sich am TAI nach Spielberger und decken die emotionale, aufgeteilt in die affektive und physiologische Dimension, sowie die kognitive Dimension der Prüfungsangst ab.

Aus methodischen Gründen wurde dabei im gesamten Fragebogen bewusst auf die Verwendung des Begriffs "Prüfungsangst" verzichtet, um mögliche Priming Effekte und damit verbundene Verzerrungen zu vermeiden. Stattdessen wurden zur Untersuchung des Einflusses von ChatGPT auf Prüfungsangst drei Aussagen formuliert, die jeweils eine der drei Dimension abbilden. Die Bewertung erfolgte jeweils auch auf einer fünfstufigen Likert-Skala.

Ein weiterer Abschnitt des Fragebogens beschäftigte sich mit möglichen Veränderungen in der Klausurvorbereitung durch den Einsatz von ChatGPT. Zudem bewerteten die Teilnehmenden, wie wichtig sie bestimmte neue Kompetenzen im Umgang mit ChatGPT einschätzen. Die Antworten erfolgten je nach Fragestellung ebenfalls auf einer fünfstufigen Skala, etwa nach Häufigkeit oder Wichtigkeit.

Die Zusammenstellung des Fragebogens orientierte sich an gängigen quantitativen Standards der empirischen Sozialforschung und wurde im Hinblick auf Verständlichkeit, inhaltliche Relevanz und Anwendungsfreundlichkeit mehrfach überarbeitet und getestet.

### **7.3.3 Datenerhebung und statistische Auswertungsverfahren**

Nach Abschluss der Befragung wurden die erhobenen Daten als Excel-Datei exportiert. Um eine statistische Auswertung zu ermöglichen, wurden die ursprünglich als Text vorliegenden Likert-Skalen, in numerische Werte (1–5) codiert. Mehrfachantworten wurden im Rahmen der Datenaufbereitung als binäre Variablen (0 = nicht gewählt, 1 = gewählt) codiert.

Die Datenaufbereitung und Analyse erfolgten mithilfe der Programme Microsoft Excel und JASP. Zunächst wurden deskriptive Statistiken berechnet (z.B. Mittelwerte, Häufigkeiten, Standardabweichungen), um grobe Tendenzen und Verteilungen der Antworten zu untersuchen. Es wurde jeweils auch der Shapiro-Wilk-Test durchgeführt. Dieser dient dazu, die Normalverteilung der jeweiligen Daten zu überprüfen.

Zusätzlich wurden inferenzstatistische Verfahren verwendet, um mögliche Zusammenhänge zwischen der Nutzung von ChatGPT und dem Erleben von Prüfungsangst zu analysieren. Das Spearmans-Korrelationsverfahren wurde verwendet, um Zusammenhänge zwischen Prüfungsangst und der Bewertung der Nutzung von ChatGPT zu untersuchen. Um die Unterschiede zwischen mehreren Semestergruppen zu untersuchen, wurde der Kruskal-Wallis H-Test eingesetzt. Zusätzlich wurde mit Hilfe des Mann-Whitney-U-Tests geschlechterspezifische Unterschiede analysiert. Alle Verfahren wurden verwendet, da die Daten überwiegend nicht normalverteilt waren.

### **7.3.4 Gütekriterien**

Um die Qualität dieser Arbeit sicherzustellen, wurden die drei zentralen Gütekriterien Objektivität, Reliabilität und Validität berücksichtigt, da sie eine entscheidende Rolle in der empirischen Forschung spielen (Rammstedt, 2010, S.239).

Objektivität bedeutet, dass die Ergebnisse unabhängig von der durchführenden Person erfolgen. In dieser Untersuchung wurde sie durch den Einsatz eines standardisierten Online-Fragebogens sichergestellt. Dabei erhielten alle Teilnehmenden identische Fragen in derselben Reihenfolge. Zusätzlich stellte die anonyme Durchführung der Befragung über Microsoft Forms sicher, dass persönliche Einflüsse das Antwortverhalten nicht beeinflussen konnten.

Reliabilität ist gegeben, wenn Wiederholungen von Untersuchungen zu zuverlässigen und konsistenten Ergebnissen führen. Zur Steigerung der Reliabilität wurden ausschließlich geschlossene Fragen mit klar definierten Skalen, wie zum Beispiel der fünfstufigen Likert-Skala, verwendet. Zudem wurde das theoretische Konstrukt der Prüfungsangst über mehrere Dimensionen abgebildet.

Validität bezieht sich darauf, ob tatsächlich das gemessen wurde, was gemessen werden sollte. Um diese zu gewährleisten, orientierte sich die Erhebung an etablierten Messinstrumenten wie dem TAI. Die Aussagen wurden so formuliert, dass sie das theoretische Konstrukt möglichst präzise widerspiegeln. Der Fragebogen wurde zudem vorab intern mehrfach auf Verständlichkeit, inhaltliche Relevanz und Anwendungsfreundlichkeit überprüft.

Die Berücksichtigung dieser Gütekriterien bildet somit eine zentrale Grundlage für die Aussagekraft und wissenschaftliche Belastbarkeit der vorliegenden Untersuchung.

## 7.4 Ergebnisse der Datenanalyse

In diesem Kapitel werden die gesammelten Daten ausgewertet und analysiert. Zur Überprüfung von möglichen Zusammenhängen zwischen der Nutzung von ChatGPT und verschiedenen Dimensionen der Prüfungsangst werden inferenzstatistische Analysen durchgeführt.

### 7.4.1 Deskriptive Statistiken

Für die deskriptive Analyse wurden nur die Fragen berücksichtigt, die direkt zur Beantwortung der Forschungsfrage und den Hypothesen beitragen.

Die Auswertung der Frage, wofür ChatGPT am häufigsten bei der Klausurvorbereitung genutzt wird, zeigt klare Tendenzen unter den Teilnehmenden. Am häufigsten gaben die Befragten an, ChatGPT zum Erklären von schwierigen Inhalten zu nutzen, gefolgt vom Üben mit Beispielaufgaben und dem Erstellen von Zusammenfassungen. Weniger häufig wurde angegeben, ChatGPT auf eine andere Weise zu nutzen oder gar keine KI-Tools zur Klausurvorbereitung einzusetzen (siehe Abbildung 2).

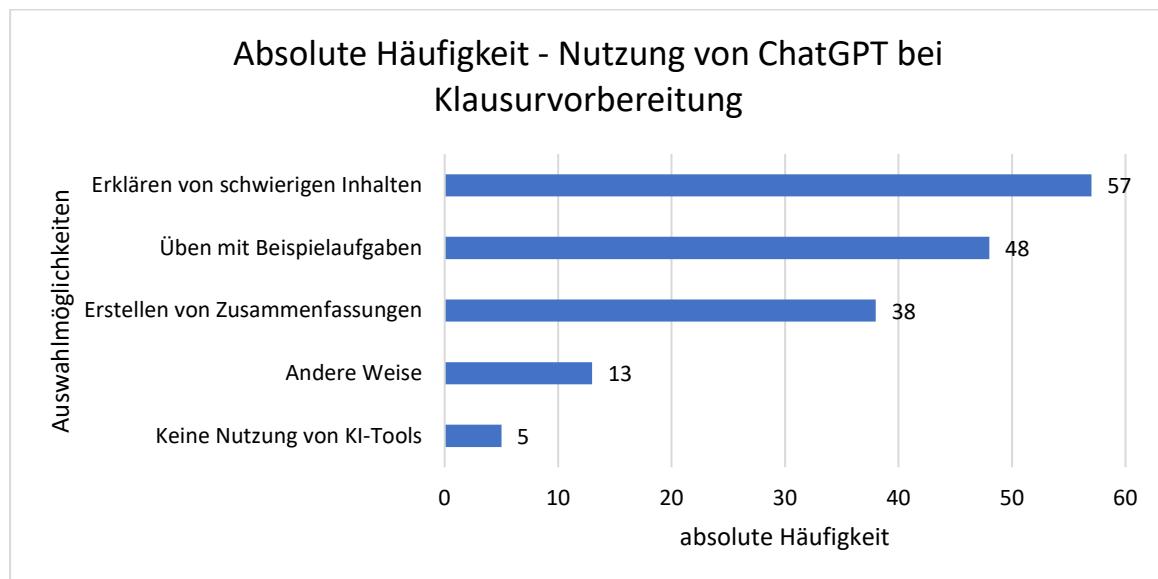


Abbildung 7.2: Absolute Häufigkeit der Nutzungsalternativen von ChatGPT bei Klausurvorbereitung

Mithilfe des Shapiro-Wilk-Tests konnte festgestellt werden, dass die Verteilungen aller Antwortoptionen signifikant von der Normalverteilung abweichen (siehe Tabelle 1). Da es sich jedoch um binäre Variablen handelt, ist dieses Ergebnis erwartbar gewesen.

	Erklären von schwierigen Inhalten	Erstellen von Zusammenfassungen	Üben mit Beispiel-aufgaben	Andere Weise	Keine Nutzung von KI-Tools
Mittelwert	0,864	0,576	0,727	0,197	0,076
Standard-abweichung	0,346	0,498	0,449	0,401	0,267
Shapiro-Wilk	0,406	0,628	0,557	0,487	0,292
P-Wert (Shapiro Wilk)	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

Tabelle 7.1: Deskriptive Statistiken zur Nutzungsalternativen von ChatGPT bei Klausurvorbereitung  
(Quelle: JASP)

In der nächsten Frage sollten die Teilnehmenden angeben, auf welche klassische Lernmethoden sie bei der Klausurvorbereitung zurückgreifen würden, wenn es kein ChatGPT geben würde. Studierende würden besonders häufig auf das Lesen von Mitschriften, das Arbeiten mit Lehrbüchern und Skripten sowie das Erstellen eigener Zusammenfassungen zurückgreifen. Gemeinsames Lernen wurde im Vergleich seltener gewählt (siehe Abbildung 3).

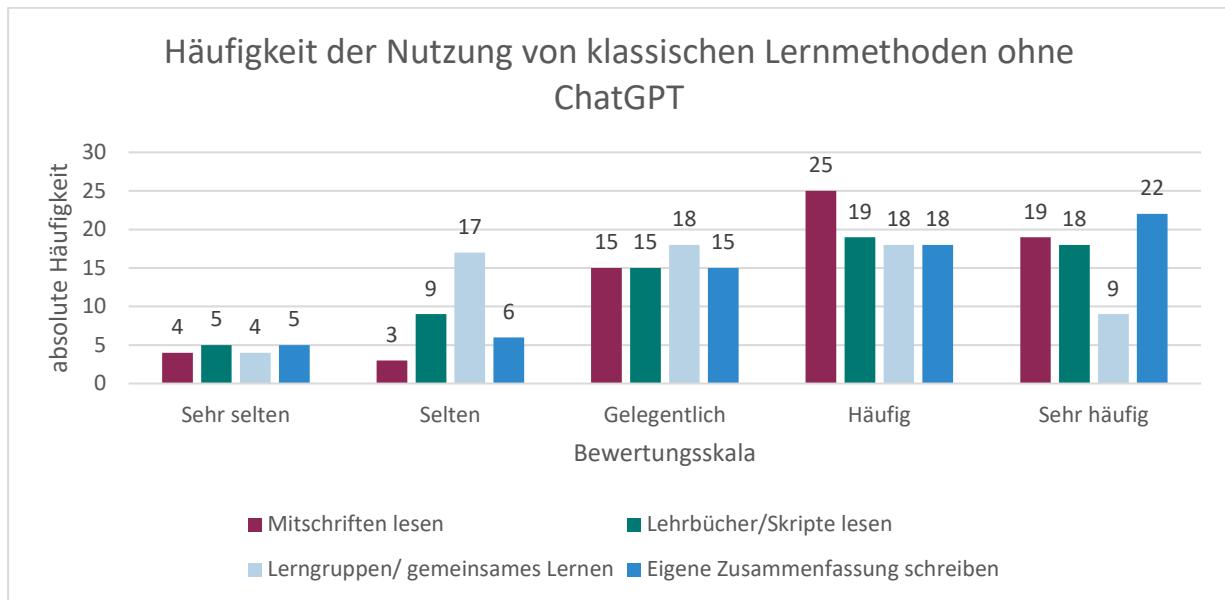


Abbildung 7.3: Häufigkeit der Nutzung von klassischen Lernmethoden ohne ChatGPT

Ähnlich wie bei den anderen Statistiken könnte auch hier mit Hilfe des Shapiro-Wilk-Test eine signifikante Abweichung von der Normalverteilung festgestellt werden, wie in Tabelle 2 ersichtlich wird.

	Mitschriften lesen	Lehrbücher / Skripte lesen	Lerngruppen / gemeinsames Lernen	Eigene Zusammenfassung
Mittelwert	3,788	3,545	3,167	3,697
Standardabweichung	1,103	1,243	1,145	1,240
Shapiro-Wilk	0,849	0,884	0,911	0,860
P-Wert Shapiro Wilk	< 0,001	< 0,001	< 0,001	< 0,001

Tabelle 7.2: Deskriptive Statistiken für klassische Lernmethoden bei der Klausurvorbereitung ohne ChatGPT (Quelle: JASP)

Anschließend sollten die Teilnehmenden bewerten, wie wichtig verschiedenen Fähigkeiten für eine sinnvolle Nutzung von ChatGPT bei der Klausurvorbereitung sind.

Die Beurteilung erfolgte mittels einer fünfstufigen Likert-Skala, wobei 1 für „unwichtig“ und 5 für „sehr wichtig“ stand.

Die Betrachtung der genauen Häufigkeitsverteilungen (Abbildung 4) zeigen, dass von den befragten Studierenden insbesondere die Fähigkeiten Ergebnisse kritisch zu bewerten, Prompts richtig zu formulieren und Inhalte in eigenen Worten wiederzugeben von als besonders wichtig eingestuft wurden.

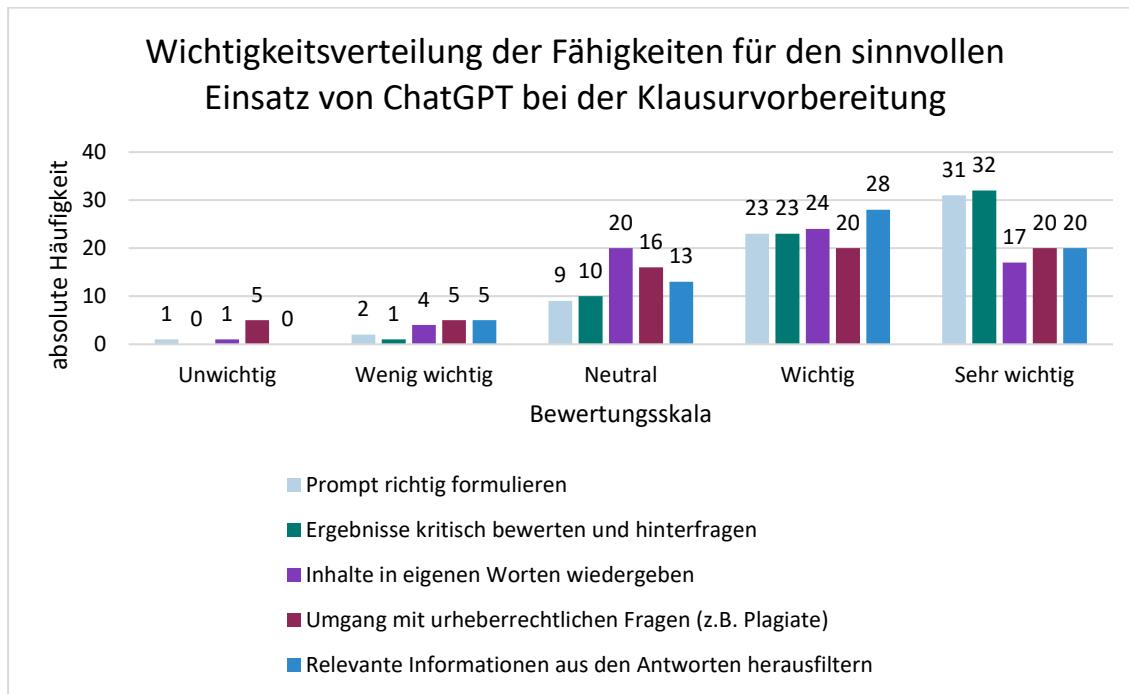


Abbildung 7.4: Häufigkeitsverteilungen der Bewertungen wichtiger Fähigkeiten für die Nutzung von ChatGPT bei der Klausurvorbereitung

Die Aspekte Umgang mit urheberrechtlichen Fragen und relevante Informationen aus den Antworten herausfiltern wurden als weniger wichtig bewertet (siehe Tabelle 3).

	Mittelwert	Standard-abweichung	Shapiro-Wilk-Test	P-Wert (Shapiro-Wilk-Test)
Prompt richtig formulieren	4,23	0,91	0,78	< ,001
Ergebnisse kritisch bewerten und hinter-fragen	4,30	0,78	0,78	< ,001
Inhalte in eigenen Worten wieder-geben	3,79	0,95	0,88	< ,001
Umgang mit urheber-rechtlichen Fragen (z.B. Plagiate)	3,68	1,20	0,86	< ,001
Relevante Informationen herausfiltern	3,95	0,90	0,85	< ,001

Tabelle 7.3: Deskriptive Statistiken zur Bewertung relevanter Fähigkeiten für die Nutzung von ChatGPT  
(Quelle: JASP)

Der Shapiro-Wilk-Test zeigte, wie zuvor, dass alle Fähigkeiten signifikante Abweichungen von der Normalverteilung aufwiesen (siehe Tabelle 3).

In der nächsten Fragestellung wurden vier Aussagen zu den einzelnen Dimensionen zur Messung der allgemeinen Prüfungsangst erfasst. Sie werden jeweils mit „In Prüfungssituationen ...“ eingeleitet. Die Zustimmungswerte sind in der Häufigkeitsverteilung (Abbildung 4) ersichtlich. Dabei zeigte sich, dass die kognitive Aussage „mache ich mir Gedanken, ob ich rechtzeitig fertig werde“ und die affektive Aussage „empfinde ich eine innere Unruhe“ am häufigsten zustimmend bewertet wurden, gefolgt von der physiologischen Dimension „fühle ich mich angespannt und verkrampt“ und der weiteren affektiven Aussage „habe ich ein komisches Gefühl im Bauch“.

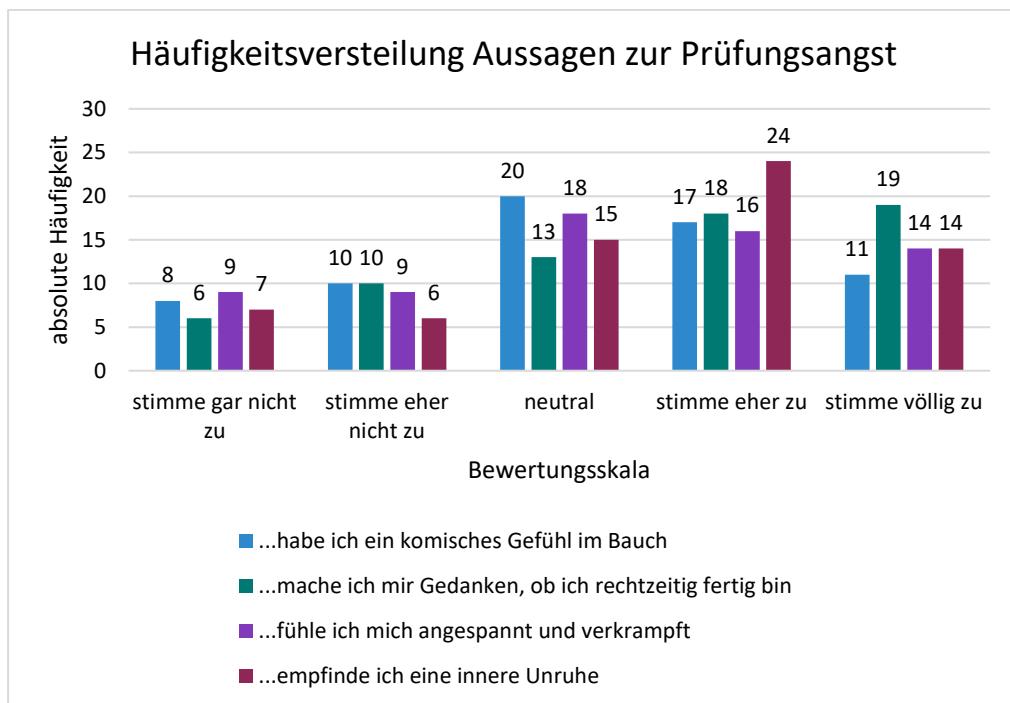


Abbildung 7.5: Häufigkeitsverteilung Aussagen zur Prüfungsangst

Alle vier Aussagen weisen wie zuvor in den Shapiro-Wilk-Tests signifikante Abweichungen der Normalverteilung auf (siehe Tabelle 4).

In Prüfungssituationen...	Mittelwert	Standard-abweichung	Shapiro-Wilk-Test	P-Wert (Shapiro-Wilk-Test)
...habe ich ein komisches Gefühl im Bauch. (affektiv)	3,20	1,24	0,91	< ,001
...mache ich mir Gedanken, ob ich rechtzeitig fertig bin. (kognitiv)	3,52	1,30	0,88	< ,001
...fühle ich mich angespannt und verkrampt. (physiologisch)	3,26	1,32	0,90	< ,001
...empfinde ich eine innere Unruhe. (affektiv)	3,48	1,23	0,87	< ,001

Tabelle 7.4: Deskriptive Statistiken Aussagen zu unterschiedlichen Dimensionen der Prüfungsangst  
(Quelle: JASP)

Zur weiteren Auswertung in Bezug auf die Prüfungsangst wurde ein Gesamt-Score der Prüfungsangst erstellt. Der Gesamtwert für die individuelle Ausprägung der Prüfungsangst wurde durch Bildung eines Mittelwerts aus den vier Aussagen berechnet. In der Stichprobe liegt der durchschnittliche Prüfungsangst-Score bei 3,36 auf der fünfstufigen Bewertungsskala. Wie bereits vorhin bei der Einzelbetrachtung wurde auch hier eine signifikante Abweichung von der Normalverteilung festgestellt mit  $p = 0,023$  (siehe Tabelle 5).

	Gesamt-Score Prüfungsangst
Mittelwert	3,36
Standardabweichung	1,03
Shapiro-Wilk	0,96
P-Wert Shapiro Wilk	0,02

Tabelle 7.5: Deskriptive Statistiken Gesamt Prüfungsangst-Score (Quelle: JASP)

Mit der nächsten Fragestellung wurde mittels dreier Aussagen für jeweils eine Dimension der Prüfungsangst die Wahrnehmung der Studierenden bezüglich ChatGPT im Prüfungskontext bei der Klausurvorbereitung erfasst. Die zugehörigen Häufigkeitsverteilungen in Abbildung 6 verdeutlichen die Tendenzen. Dabei ist zu beachten, dass die Aussage zur affektiven Komponente negativ und die beiden anderen positiv formuliert sind.

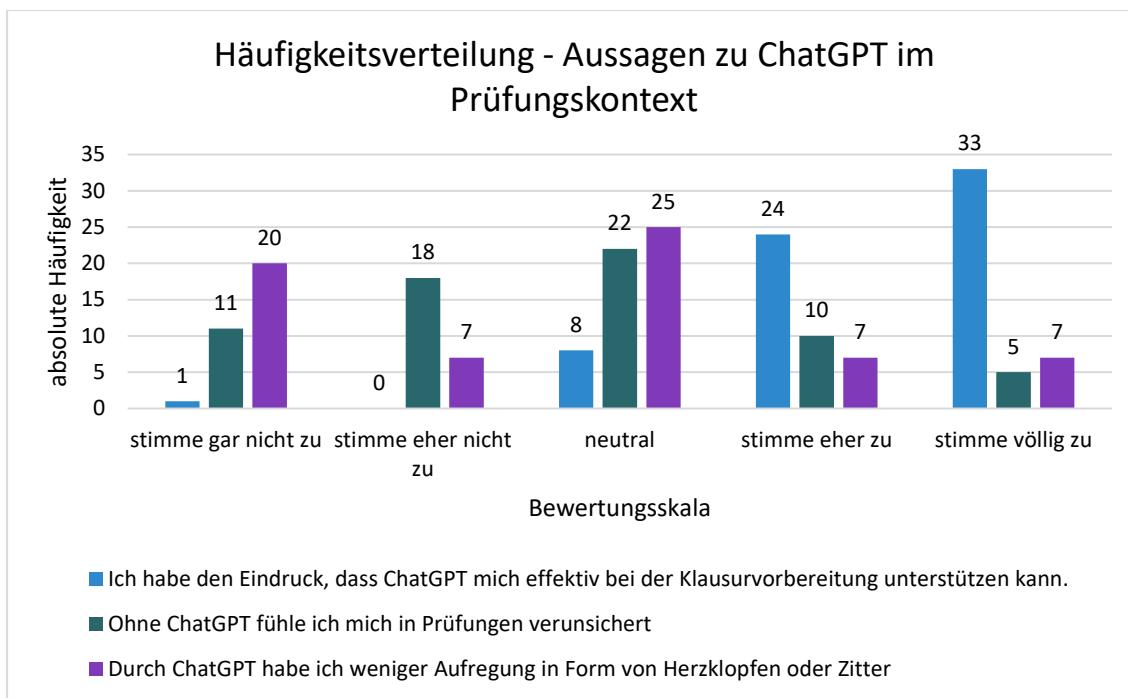


Abbildung 7.6: Häufigkeitsverteilung - Aussagen zu ChatGPT im Prüfungskontext

Die kognitive Aussage „Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“ wurde sehr positiv bewertet. Die affektive Komponente „Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“ wies eine geringere Zustimmung auf. Ebenso wurde ein direkter Einfluss auf die physiologische Dimension „Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern“ gering eingeschätzt.

Auch hier weisen alle drei Aussagen signifikante statistische Abweichungen von der Normalverteilung auf (siehe Tabelle 6).

	„Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“ (kognitiv)	„Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“ (affektiv)	„Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern.“ (physiologisch)
Mittelwert	4,33	2,70	2,61
Standardabweichung	0,81	1,15	1,31
Shapiro-Wilk	0,75	0,91	0,87
P-Wert Shapiro Wilk	< ,001	< ,001	< ,001

Tabelle 7.6: Deskriptive Statistiken - Aussagen zu ChatGPT im Prüfungskontext (Quelle: JASP)

#### 7.4.2 Gruppenvergleiche

Zusätzlich wurden Gruppenvergleiche durchgeführt, um potenzielle Unterschiede in den zentralen Variablen auf Grundlage demografischer Merkmale zu untersuchen.

Zur Überprüfung von geschlechtsspezifischen Unterschieden im Hinblick auf Prüfungsangst und die subjektive Wahrnehmung von ChatGPT in der Klausurvorbereitung wurde der Mann-Whitney-U-Test durchgeführt. Dieses Verfahren wurde gewählt, da die zugrunde liegenden Daten nicht normalverteilt sind und auf Likert-Skalen beruhen. Studierende mit der Angabe „divers“ wurden aufgrund einer zu geringen Anzahl ( $n = 1$ ) vom Vergleich ausgeschlossen, um Verzerrungen zu vermeiden.

Zur Analyse wurden Mittelwerte und Standardabweichungen der Geschlechter sowie Mann-Whitney-U-Tests für die einzelnen Aussagen berechnet (siehe Tabelle 7 und 8).

	Gruppe	Mittelwert	Standard-abweichung	Mann-Whitney-U	p-Wert	Effekt-stärke
„....habe ich ein komisches Gefühl im Bauch.“	M	2,96	1,3	312,5	0,045	-0,31
„....habe ich ein komisches Gefühl im Bauch.“	W	3,65	0,93			
„....mache ich mir Gedanken, ob ich rechtzeitig fertig bin.“	M	3,38	1,37	350,5	0,147	-0,22
„....mache ich mir Gedanken, ob ich rechtzeitig fertig bin.“	W	3,95	0,95			
„....fühle ich mich angespannt und verkrampt.“	M	3,4	1,32	304,5	0,034	-0,32
„....fühle ich mich angespannt und verkrampt.“	W	3,01	1,32			
„....empfinde ich eine innere Unruhe.“	M	3,4	1,27	368,5	0,232	-0,18
„....empfinde ich eine innere Unruhe.“	W	3,8	1,01			

Tabelle 7.7: Gruppenvergleiche zwischen männlichen und weiblichen Teilnehmenden hinsichtlich Prüfungsangst

	Gruppe	Mittel-wert	Standard-abweichung	Mann-Whitney-U	p-Wert	Effekt-stärke
„Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“	M	4,49	0,63	552	0,11	0,227
„Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“	W	4,15	0,81			
„Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“	M	2,76	1,15	488,5	0,58	0,086
„Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“	W	2,6	1,19			
„Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern“	M	2,71	1,36	497	0,49	0,1
„Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern“	W	2,45	1,19			

Tabelle 7.8: Gruppenvergleiche zwischen männlichen und weiblichen Teilnehmenden ChatGPT-bezogenen Aussagen

Zum einen berichteten weibliche Teilnehmende signifikant häufiger in Prüfungssituationen ein „komisches Gefühl im Bauch“ zu haben ( $U = 312,5$ ;  $p = 0,045$ ;  $r = -0,31$ ). Zum anderen fühlen sich weibliche Teilnehmende in Prüfungssituationen signifikant häufiger „angespannt und verkrampt“, als männliche Teilnehmende ( $U = 304,5$ ;  $p = 0,034$ ;  $r = -0,32$ ). In beiden Fällen weiß die Effektstärke auf einen mittleren Effekt hin.

Bei den übrigen Aussagen zur Prüfungsangst konnten keine signifikanten Unterschiede zwischen den Geschlechtern festgestellt werden.

Auch bei den Aussagen bezüglich der ChatGPT-Nutzung traten keine geschlechterspezifischen signifikanten Unterschiede auf. Die Aussage „Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann“ zeigte zwar eine moderate, aber nicht signifikanten Tendenz zugunsten der männlichen Teilnehmenden ( $U = 552$ ;  $p = 0,11$ ;  $r = 0,227$ ). Ähnliches konnte auch für die Aussage „Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern“ beobachtet werden ( $U = 497$ ;  $p = 0,49$ ;  $r = 0,10$ ).

Insgesamt deuten die Ergebnisse drauf hin, dass sich insbesondere im affektiven und physiologischen Bereich der Prüfungsangst geschlechterspezifische Unterschiede zeigen. Da beide Dimensionen der emotionalen Komponente zugeordnet werden können, lässt sich hier ein möglicher Zusammenhang vermuten. Jedoch konnten hinsichtlich der Bewertung der ChatGPT-Nutzung im Prüfungskontext keine signifikanten Unterschiede zwischen den Geschlechtern festgestellt werden.

Für die Vergleiche zwischen der Semesterzugehörigkeit der befragten Studierenden und den einzelnen Aussagen zur Prüfungsangst sowie den Wahrnehmungen zu ChatGPT im Prüfungskontext wurde aufgrund der nicht normal verteilten Daten und der mehr als zwei Vergleichsgruppen ein Kruskal-Wallis-H-Test durchgeführt.

Die durchgeführten Tests für die einzelnen Aussagen zur Prüfungsangst in Bezug auf die Semesterzugehörigkeit ergaben durchweg p-Werte, die das Signifikanzniveau von 0,05 überschritten (siehe Tabelle 9). Somit kann aus den vorliegenden Daten kein statistisch signifikanter Zusammenhang zwischen dem Semester und der Ausprägung der einzelnen Prüfungsangstdimensionen abgeleitet werden.

In Prüfungssituationen...	...habe ich ein komisches Gefühl im Bauch.	...mache ich mir Gedanken, ob ich rechtzeitig fertig bin.	...fühle ich mich verkrampt und angespannt.	...empfinde ich eine innere Unruhe.
Statistik	2,08	3,57	1,58	2,68
df	3	3	3	3
p	0,56	0,31	0,66	0,44

Tabelle 7.9: Kruskal-Wallis-Test: Vergleich von Semester und Aussagen zur Prüfungsangst (Quelle:JASP)

Bezüglich der Aussagen zu ChatGPT im Prüfungskontext zeigten sich für die Aussagen „Ohne ChatGPT fühle ich mich in Prüfungen verunsichert“ und „Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern“ ebenfalls keine signifikanten Unterschiede in Abhängigkeit der Semesterzugehörigkeit. Für die Aussage „Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“ zeigte der Kruskal-Wallis H-Test hingegen eine signifikante Abweichung mit einem p-Wert von 0,03 (siehe Tabelle 10). Um dies weiter zu beleuchten, wurde ein Dunns-Post-hoc-Test durchgeführt.

	„Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“	„Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“	„Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern.“
Statistik	8,83	7,15	4,27
df	3	3	3
p	0,03	0,07	0,23

Tabelle 7.10: Kruskal-Wallis-Test Vergleich von Semester und Aussagen zu ChatGPT im Prüfungskontext (Quelle:JASP)

Um die genauen Unterschiede zwischen den Semestergruppen zu identifizieren, wurden die Holm-korrigierten p-Werte (pHolm) des Dunn-Post-hoc-Tests sowie die rangbiseriale Korrelation (rrb) als Maß für die Effektstärke berücksichtigt.

Es zeigte sich, dass die Einschätzung der Effektivität von ChatGPT für die Klausurvorbereitung zwischen den Studierenden aus dem 2. Semester und den Studierenden, die „Über 8“ gewählt haben, signifikant voneinander abweicht ( $p_{Holm} = 0,04$ , starker Effekt mit  $rrb = 0,71$ ). Die Studierenden des 2. Semesters schätzten die Effektivität hierbei höher ein. Auch im Vergleich zwischen dem 6. Semester und der Semestergruppe „Über 8“ ( $p_{Holm} = 0,04$ , sehr starker Effekt mit  $rrb = 0,86$ ), bewerteten Studierende im 6. Semester die Effektivität erneut höher (siehe Tabelle 11).

Semestergruppenvergleiche	$z$	$W_i$	$W_j$	$r_{rb}$	$p$	$p_{Bonf}$	$p_{Holm}$
2 versus 4	1,15	37,83	32,09	0,18	0,25	1,00	0,52
2 versus 6	-0,52	37,83	41,86	0,10	0,60	1,00	0,60
2 versus Über 8	2,64	37,83	12,38	0,71	$8.27 \times 10^{-3}$	0,05	0,04
4 versus 6	-1,36	32,09	41,86	0,31	0,17	1,00	0,52
4 versus Über 8	2,15	32,09	12,38	0,64	0,03	0,19	0,13
6 versus Über 8	2,70	41,86	12,38	0,86	$6.99 \times 10^{-3}$	0,04	0,04

Tabelle 7.11: Dunns-Post-hoc-Test von Semester und Aussagen zu ChatGPT im Prüfungskontext (Quelle: JASP)

#### 7.4.3 Zentrale Befunde

Im Folgenden werden mithilfe von Korrelationsanalysen die wesentlichen Zusammenhänge zwischen den Variablen aufgezeigt, die für die Beantwortung der Forschungsfrage und der Hypothese maßgeblich sind. Ziel dieser Analyse war es einen Zusammenhang zwischen der Nutzung von ChatGPT und dem Erleben von Prüfungsangst herzustellen.

Zur Untersuchung des Zusammenhangs zwischen der Häufigkeit der Nutzung generativer KI bei der Klausurvorbereitung und der Ausprägung der Gesamtprüfungsangst der Befragten wurde eine Korrelationsanalyse durchgeführt. Da beide Variablen keine Normalverteilung aufwiesen, wurde der Spearman-Rangkorrelationskoeffizient eingesetzt, der auf ordinale Daten ausgelegt ist.

Die Auswertung ergab einen Spearman-Rho-Wert von -0,13 was auf einen schwachen negativen Zusammenhang hindeutet (siehe Tabelle 12). Ein negativer Zusammenhang bedeutet in diesem Fall, dass eine höhere Nutzung von ChatGPT tendenziell mit einer leicht reduzierten Prüfungsangst einhergeht.

Häufigkeit der Nutzung von ChatGPT im universitären Kontext		
Gesamt-Prüfungsangstscore	$\rho$ (rho)	-0,13
	p-Wert	0,31

Tabelle 7.12: Spearman Korrelation zwischen Gesamt-Prüfungsangstscore und Häufigkeit der Nutzung von ChatGPT im universitären Kontext (Quelle: JASP)

Da der beobachtete Zusammenhang keine statistische Signifikanz erreichte ( $p = 0,31$ ) ließ sich kein belastbarer Zusammenhang zwischen der Häufigkeit der ChatGPT-Nutzung im Studium und der Ausprägung der Gesamtprüfungsangst feststellen.

Zur vertiefenden Analyse wurde überprüft, inwiefern die unterschiedlichen Dimensionen der Prüfungsangst mit der Nutzung von ChatGPT in der Prüfungsvorbereitung im Zusammenhang stehen. Die Ergebnisse der Korrelationsanalyse sind in Tabelle 13 dargestellt.

Prüfungsangst Aussage (In Prüfungssituationen...)	ChatGPT-Aussage	$\rho$ (rho)	p-Wert
...habe ich ein komisches Gefühl im Bauch.	„Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“	0,119	0,341
...empfinde ich eine innere Unruhe.	„Ohne ChatGPT fühle ich mich in Prüfungen verunsichert.“	0,162	0,193
...mache ich mir Gedanken, ob ich rechtzeitig fertig werde.	„Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“	0,157	0,208
...fühle ich mich angespannt und verkrampft.	„Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern.“	0,067	0,594

Tabelle 7.13: Spearman Korrelation zwischen Prüfungsangst-Aussagen und Einschätzung zur ChatGPT-Nutzung im Prüfungskontext (Quelle: JASP)

Wie zuvor bereits erläutert, basieren alle Merkmale auf fünfstufigen Likert-Skalen. Aus diesem Grund wurde auch hier der Spearman-Rangkorrelationskoeffizient verwendet.

Zwischen der affektiven Aussage „Ich habe in Prüfungssituationen ein komisches Gefühl im Bauch“ und „Ohne ChatGPT fühle ich mich in Prüfungen verunsichert“ ergab sich ein schwacher positiver nicht signifikanter Zusammenhang ( $\rho = 0,119$ ;  $p = 0,341$ ). Ein schwächerer, jedoch ebenfalls nicht signifikanter Zusammenhang bestand zwischen der zweiten affektiven Aussage „In Prüfungssituationen empfinde ich eine innere Unruhe“ und derselben ChatGPT bezogenen Aussagen ( $\rho = 0,162$ ;  $p = 0,193$ ).

Auch bei der kognitiven Aussage „Ich mache mir Gedanken, ob ich rechtzeitig fertig werde“ und der positiv formulierten Einschätzung „Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann“ konnte ein moderater, aber nicht signifikanter Zusammenhang festgestellt werden. ( $\rho = 0,157$ ;  $p = 0,208$ ). Das deutet daraufhin, dass Studierende mir stärkerer kognitiven Prüfungsangst ChatGPT einen höheren Nutzen zusprechen.

Im physiologischen Bereich zeigte sich zwischen „In Prüfungssituationen fühle ich mich angespannt und verkrampft“ und der positiven Einschätzung „Durch ChatGPT habe ich weniger Aufregung in Form von Herzklopfen oder Zittern“ ein sehr schwacher und nicht signifikanter positiver Zusammenhang ( $\rho = 0,067$ ;  $p = 0,594$ ).

Insgesamt deuten die Ergebnisse auf leichte positive Zusammenhänge zwischen den verschiedenen Aspekten der Prüfungsangst und einer positiven Einschätzung zur Wirksamkeit von ChatGPT hin. Jedoch ist keine Korrelation statistisch signifikant, weswegen keine belastbaren Aussagen über einen systematischen Zusammenhang möglich sind.

## **7.5 Zwischen Erkenntnis und Bedeutung**

Im folgenden Abschnitt werden die zentralen Ergebnisse der Studie im Hinblick auf die Forschungsfrage interpretiert und in den bestehenden theoretischen sowie empirischen Kontext eingeordnet. Außerdem werden die praktischen Implikationen herausgearbeitet sowie Perspektiven für die zukünftige Forschung diskutiert.

### **7.5.1 Interpretation der Ergebnisse im Kontext der Forschungsfrage**

Das Ziel der Arbeit war es zu untersuchen, ob die Nutzung von ChatGPT bei der Klausurvorbereitung das Erleben von Prüfungsangst beeinflusst. Dazu wurde die Nutzungshäufigkeit von ChatGPT und verschiedene Dimensionen der Prüfungsangst erfasst und mittels Korrelationsanalyse in Beziehung gesetzt. Zwischen der allgemeinen Nutzung generativer KI und dem Gesamtwert der Prüfungsangst zeigte sich ein schwacher negativer, aber nicht signifikanter Zusammenhang ( $p = -0,13$ ;  $p = 0,31$ ). Dies deutet darauf hin, dass eine häufigere Nutzung von ChatGPT tendenziell zu geringerer Prüfungsangst führen kann, wobei dieser Effekt nicht als systematischer Zusammenhang interpretierbar ist.

Darauf aufbauend wurden die affektive, kognitive und physiologische Dimension gesondert analysiert. Hierbei konnten ebenfalls schwache, positive, jedoch nicht signifikante Zusammenhänge zu entsprechenden Einschätzungen zur Nutzung von ChatGPT beobachtet werden. Am stärksten war der Zusammenhang im affektiven Bereich zwischen den Aussagen „In Prüfungssituationen empfinde ich eine innere Unruhe“ und „Ohne ChatGPT fühle ich mich in Prüfungen verunsichert“ ( $p = 0,162$ ). Da auch hier keine statistische Signifikanz vorliegt, kann nur vermutet werden, dass Studierende mit stärker ausgeprägter Prüfungsangst ChatGPT als entlastend empfinden, ohne dass daraus belastbare Aussagen abgeleitet werden können.

Zusätzlich wurden geschlechterspezifische Unterschiede bei der Prüfungsangst und der Bewertung von ChatGPT analysiert. Bei den affektiven und physiologischen Aussagen berichteten weibliche Teilnehmende signifikant häufiger Symptome wie ein „komisches Gefühl im Bauch“ und sich „angespannt und verkrampt“ zu fühlen. Dies spricht dafür, dass Frauen emotional und physiologisch stärker auf Prüfungssituationen reagieren (Núñez-Peña et al., 2016, S.154). In Bezug auf die Bewertung von ChatGPT zeigten sich keine signifikanten Geschlechterunterschiede.

Darüber hinaus wurde der Einfluss der Semesterzugehörigkeit untersucht. Die Kruskal-Wallis H-Tests zu Aussagen der Prüfungsangst ergaben keine signifikanten Unterschiede zwischen den Gruppen und somit auch keinen systematischen Zusammenhang zwischen dem Studienfortschritt und dem Erleben von Prüfungsangst.

Hinsichtlich der Einschätzung von ChatGPT ergab sich bei der Aussage „Ich habe den Eindruck, dass ChatGPT mich effektiv bei der Klausurvorbereitung unterstützen kann.“ ein signifikanter Unterschied ( $p = 0,03$ ). Studierende im 2. und 6. Semester bewerteten die Effektivität deutlich höher als Studierende aus höheren Semestern („über 8“), mit einem starken Effekt ( $rrb = 0,71$  bzw.  $0,86$ ). Das deutet daraufhin, dass insbesondere Studierende in früheren oder mittleren Semestern mehr Potential in ChatGPT sehen. Mögliche Gründe dafür könnten in den altersbedingten Unterschieden im Umgang mit digitalen Technologien liegen. Häufig sind jüngere Studierende in niedrigeren Semestern offener für digitale Hilfsmittel wie KI-Tools, während ältere Studierende in höheren Semestern diesen Technologien eher skeptisch gegenüberstehen. Diese Annahme sollte in weiterführenden Studien systematisch überprüft werden.

Insgesamt deuten die Ergebnisse auf mögliche Zusammenhänge zwischen Prüfungsangst und der subjektiven Bewertung von ChatGPT hin. Aufgrund fehlender Signifikanz sind jedoch keine

belastbaren Aussagen möglich. Die Ergebnisse liefern erste Anhaltspunkte für weiterführende Forschung.

### **7.5.2 Einordnung in den theoretischen und empirischen Forschungsstand**

Die Ergebnisse dieser Untersuchung fügen sich gut in bestehende Modelle empirische Studien ein. Theoretisch basiert die Untersuchung auf einem mehrdimensionalen Konstrukt von Prüfungsangst (Liebert & Morris, 1967; Spielberger, 1980; Stöber, 2004), dass kognitive, affektive und physiologische Komponenten umfasst. Diese wurden bei der Fragebogenerstellung berücksichtigt, um eine Vergleichbarkeit mit bestehenden Modellen herzustellen.

Zudem zeigen Studien, dass Selbstwirksamkeit, Leistungsdruck oder externe Rahmenbedingungen die Intensität von Prüfungsangst beeinflussen können (Zeidner, 1998; Maier et al., 2021; Zheng, 2023). Die Ergebnisse der Untersuchung deuten darauf hin, dass der Einsatz von ChatGPT eine unterstützende Funktion in der Klausurvorbereitung übernehmen kann. Dies äußert sich in einer insgesamt positiven Wirksamkeitswahrnehmung und einer geringen emotionalen Belastung.

Im Vergleich zum empirischen Forschungsstand zeigen sich Parallelen zur Gao (2024), der ebenfalls einen positiven Einfluss von ChatGPT auf Prüfungsangst und emotionale Reaktionen beobachtete. Gleichzeitig bestätigen die Ergebnisse die Befunde von Ali et al. (2025), die keinen signifikanten kurzfristigen Zusammenhang zwischen ChatGPT- Nutzung und Prüfungsangst nachweisen konnten. Dies verdeutlicht, dass die Wirkung generativer KI differenzierter zu betrachten ist, da sie je nach Dimension der Prüfungsangst unterschiedlich ausfallen kann.

Darüber hinaus stützen die Ergebnisse die Annahme, dass neue Kompetenzen notwendig sind, um ChatGPT effektiv für die Prüfungsvorbereitung zu nutzen (Schmohl et al., 2023; Alasadi & Baiz, 2023; Educational Technology Journal, 2024).

Die vorliegende Arbeit leistet somit einen Beitrag zur Schließung der Forschungslücke, indem sie den Zusammenhang zwischen KI-Nutzung und Prüfungsangst sowie deren emotionale, kognitive und physiologische Aspekte untersucht.

### **7.5.3 Praktische Implikationen und Forschungsperspektiven**

Die Erkenntnisse der Studie zeigen zwar kaum statistisch signifikanten Zusammenhänge, bieten aber erste Anhaltspunkte für die Praxis und künftige Forschung.

Hochschulen sollten erkennen, dass der bloße Einsatz generativen KI-Tools nicht ausreicht, um die Prüfungsangst zu reduzieren. Unterstützungsangebote zur Lernstrategie und Prüfungsbewältigung sollten weiterhin im Mittelpunkt stehen, wobei verstärkt auf den gezielten Einsatz generativer KI zur Entlastung Studierender geachtet werden sollte. Dabei könnten Hochschulen Studierende dabei unterstützen, den Nutzen solcher Tools kritisch zu reflektieren und zentrale Kompetenzen für den verantwortungsvollen Einsatz mit ChatGPT zu entwickeln. Aus der Sicht der Befragten sind solche Fähigkeiten entscheidend, um ChatGPT sinnvoll in der Klausurvorbereitung einzusetzen.

Um die Auswirkungen generativer KI auf Prüfungsangst besser zu untersuchen, sind vielfältigere Stichproben sowie Langzeitstudien notwendig. Einblicke in die Nutzungserfahrungen der Studierenden könnten durch weitere empirische Ansätze gewonnen werden. Zudem sollten weitere Einflussfaktoren berücksichtigt werden, um differenziertere Erkenntnisse und praxisnahe Empfehlungen ableiten zu können. Insgesamt bietet die Arbeit eine fundierte Grundlage für die weitere Auseinandersetzung mit den Auswirkungen generativer KI im Hochschulbereich.

## **7.6 Vom Ausgangspunkt zu den zentralen Tendenzen**

In diesem letzten Kapitel werden die zentralen Ergebnisse der Studie, vor allem im Hinblick auf die Forschungsfrage und Hypothesen, zusammengefasst.

### **7.6.1 Zusammenfassung der wichtigsten Erkenntnisse**

Ziel der Untersuchung war es, den Zusammenhang zwischen der Nutzung generativer KI, insbesondere ChatGPT, und dem Erleben von Prüfungsangst bei Studierenden der Wirtschaftsinformatik an der Frankfurt University of Applied Sciences zu analysieren. Die deskriptiven Ergebnisse zeigen, dass viele Studierenden ChatGPT regelmäßig zur Klausurvorbereitung nutzt, insbesondere um sich schwierige Inhalte erklären und Zusammenfassungen erstellen zu lassen. Gleichzeitig berichten viele typischen Symptomen der Prüfungsangst, wie Anspannung, innere Unruhe und kognitive Belastung.

Die inferenzstatistischen Analysen ergaben insgesamt nur schwache, nicht signifikante Zusammenhänge zwischen der Nutzung von ChatGPT und dem Erleben von Prüfungsangst. Auch differenziertere Betrachtungen einzelner Prüfungsangst-Dimensionen zeigten lediglich leichte, statistisch nicht signifikante Tendenzen.

Die Gruppenvergleichen lieferten interessante Ergebnisse. Während sich keine geschlechterspezifischen Unterschiede in der Bewertung von ChatGPT zeigten, berichteten weibliche Studierende signifikant häufiger affektive und physiologische Symptome. Zudem bewerteten Studierende in früheren Semestern die Effektivität von ChatGPT tendenziell höher als solche in fortgeschrittenen Studienphasen.

### **7.6.2 Beantwortung der Forschungsfrage und Hypothesen**

Die zentrale Forschungsfrage lautete: „Welche Auswirkungen hat die Nutzung generativer KI-Tools wie ChatGPT auf die Prüfungsangst im Kontext der Klausurvorbereitung?“

Zur Beantwortung wurden drei Hypothesen formuliert und im Rahmen der Untersuchung analysiert.

Hypothese 1 ging von einem negativen Zusammenhang zwischen der Nutzung generativer KI und dem Erleben von Prüfungsangst aus. Die Ergebnisse zeigten einen schwachen, nicht signifikanten negativen Zusammenhang zwischen der ChatGPT-Nutzung und dem Gesamtwert der Prüfungsangst. Auch differenzierte Analysen auf Basis des mehrdimensionalen Prüfungsangst-Modells lieferten keine signifikanten Effekte. Somit kann Hypothese 1 nicht bestätigt werden.

Hypothese 2 untersuchte geschlechterspezifische Unterschiede in der Prüfungsangst. Weibliche Teilnehmende berichteten signifikant häufiger affektive Symptome („komisches Gefühl im Bauch“) sowie physiologische Reaktionen („angespannt und verkrampft“) zu fühlen. In der Bewertung von ChatGPT zeigten sich jedoch keine signifikanten Geschlechtsunterschiede. Hypothese 2 lässt sich somit nur teilweise bestätigen, da hinsichtlich der Einschätzung von ChatGPT keine empirische Evidenz vorliegt.

Hypothese 3 nahm Unterschiede in der Prüfungsangst je nach Studiensemester an. In Bezug auf die Prüfungsangst konnten keine signifikanten Unterschiede beobachtet werden, während die Bewertung zur Effektivität von ChatGPT deutlich zwischen den Studiensemestern unterschied. Studierende aus dem zweiten und sechsten Fachsemester bewerteten ChatGPT signifikant positiver als Studierende höherer Fachsemester. Ein Zusammenhang zur Prüfungsangst lässt sich zwar daraus nicht ableiten, jedoch stützen die Ergebnisse Hypothese 3 teilweise.

Insgesamt zeigen sich keine signifikanten Zusammenhänge zwischen der Nutzung generativer KI und reduzierter Prüfungsangst. Einzelne Tendenzen deuten jedoch auf differenzierte Wahrnehmungen hin, die in künftiger Forschung weiter untersucht werden sollten.

## 7.7 Literaturverzeichnis

- Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in education and research: Opportunities, concerns, and solutions. *Journal of Chemical Education*, 100(8), 2965–2971.
- Ali, M., Rehman, S., & Cheema, E. (2025). Impact of artificial intelligence on the academic performance and test anxiety of pharmacy students in objective structured clinical examination: A randomized controlled trial. *International Journal of Clinical Pharmacy*.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27(2), 270–295.
- Educational Technology Journal. (2024). Embracing the future of artificial intelligence in the classroom. SpringerOpen.
- Gao, S. (2024). Can artificial intelligence give a hand to open and distributed learning? A probe into the state of undergraduate students' academic emotions and test anxiety in learning via ChatGPT. *The International Review of Research in Open and Distributed Learning*, 25(3), 199–218.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20(3), 975–978.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson Education.
- Maier, A., Schaitz, C., Kröner, J., Berger, A., Keller, F., Beschoner, P., Connemann, B., & Sosic-Vasic, Z. (2021). The association between test anxiety, self-efficacy, and mental images among university students: Results from an online survey. *Frontiers in Psychology*, 12, 758432.
- Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2016). Gender differences in test anxiety and their impact on higher education students' academic achievement. *Procedia – Social and Behavioral Sciences*, 228, 154–160.
- Pekrun, R., & Götz, T. (2006). Emotionsregulation: Vom Umgang mit Prüfungsangst. In H. Mandl (Hrsg.), *Handbuch Lernstrategien* (S. 248–258). Hogrefe.
- Rammstedt, B. (2010). Reliabilität, Validität, Objektivität. In C. Wolf & H. Best (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (S. 353–364). VS Verlag für Sozialwissenschaften.
- Schmohl, T., Watanabe, A., & Schelling, K. (2023). Künstliche Intelligenz in der Hochschulbildung: Chancen und Grenzen des KI-gestützten Lernens und Lehrens. Transcript.
- Sousa, A. E., & Cardoso, P. (2025). Nutzung generativer KI durch Studierende. *Electronics*, 14(7), 1258.
- Spielberger, C. D. (1980). Test anxiety inventory: Preliminary professional manual. Consulting Psychologists Press.
- Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: Theory, assessment, and treatment. Taylor & Francis.
- Steiner, E., & Benesch, M. (2018). Der Fragebogen: Von der Forschungsidee zur SPSS-Auswertung (5. Aufl.). Facultas.
- Stöber, J. (2004). Dimensions of test anxiety: Relations to ways of coping with pre-exam anxiety and uncertainty. *Anxiety, Stress & Coping*, 17(3), 213–226.

- von Garrel, J., & Mayer, J. (2023). Artificial intelligence in studies—Use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications*, 10, Article 363.
- Zeidner, M. (1998). Test anxiety: The state of the art. Plenum Press.
- Zheng, G., Zhang, Q., & Ran, G. (2023). The association between academic stress and test anxiety in college students: The mediating role of regulatory emotional self-efficacy and the moderating role of parental expectations. *Frontiers in Psychology*, 14, Article 1008679.

## 8 KI-Tutorensysteme

### *KI-gestützte Tutoren in der Prüfungsvorbereitung sowie Absolvierung von Prüfungsleistungen*

Chris Meier

Kerem Tekin

#### **Abstract**

In der vorliegenden Arbeit wird der Einsatz von KI-gestützten Tutorensystemen (= KI Tutoren) im Hochschulkontext untersucht. Es wird das Potential dieser Systeme analysiert, um zu ermitteln, inwiefern sie traditionelle Vorgehensweisen in der Prüfungsvorbereitung und -durchführung substituieren können. KI Tutoren werden dabei als eine auf Künstlicher Intelligenz basierende, sich individuell an Lernbedürfnisse anpassende Lernsoftware definiert. Die vorliegende Untersuchung zielt darauf ab, die Akzeptanz, Nutzungsbereitschaft sowie den Zahlungs- und Zeitaufwand im Umgang mit KI-Tutoren durch eine quantitative Erhebung unter Studierenden zu analysieren. Die Forschung kommt dabei zu dem Schluss, dass Studierende KI-Tutoren in der Prüfungsvorbereitung überwiegend positiv bewerten, sie jedoch eher als Ergänzung zu traditionellen Lernformen betrachten. Dass Studierende ein hohes Potential im Einsatz von KI-Tutoren sehen, bestätigt sich auch im wahrgenommenen Mehrwert, der überwiegend als hoch eingeschätzt wird. Die Akzeptanz für KI-basierte Bewertungen ist zum Teil gegeben, während die Nutzungsbereitschaft als gering einzustufen ist. Auch die Bereitschaft der Studierenden, Zeit mit dem KI Tutor zu verbringen, ist bemerkenswert gering, wenn die hohe Nutzungsbereitschaft im Hinterkopf bleibt. Es bestehen signifikante Korrelationen zwischen der Akzeptanz von KI-Prüfungsformaten und der Bereitschaft zur KI-Bewertung, jedoch nicht mit den zeitlichen und monetären Aspekten. Der erfolgreiche Einsatz von KI Tutoren ist daher in hohem Maße von einer schrittweisen Integration unter gleichzeitiger, zielgerichteter Aufklärungsarbeit abhängig.

**Schlagwörter:** Künstliche Intelligenz (KI), KI-Tutorensysteme, Adaptive Lernsysteme, Hochschulbildung, Prüfungsvorbereitung, Studierendenakzeptanz, Lernplattformen, Zahlungsbereitschaft, KI-basierte Leistungsbewertung.

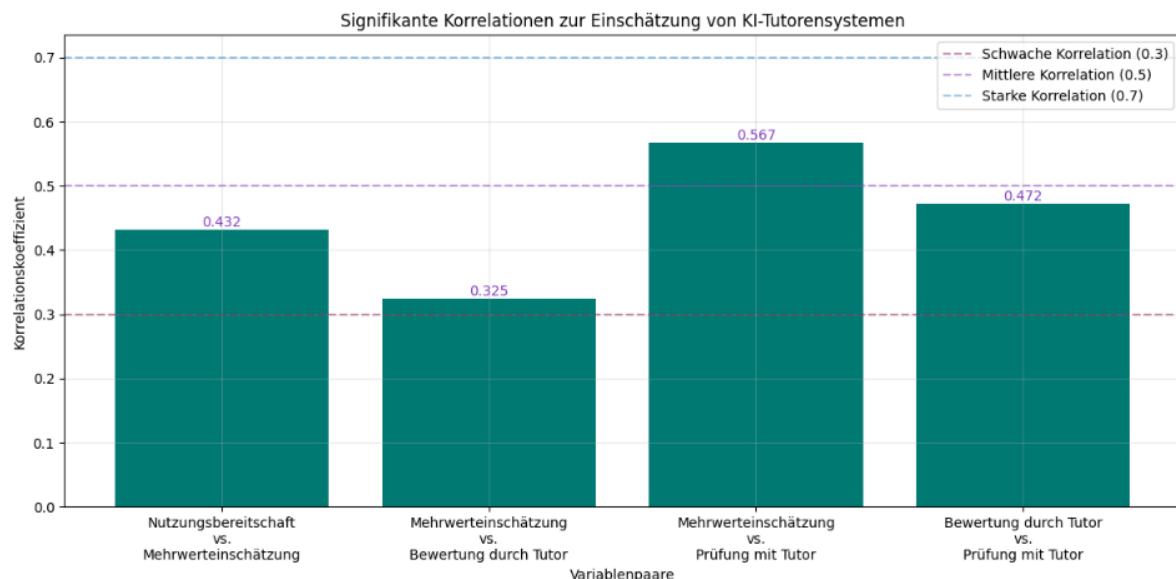


Abbildung 8.1: Zusammenhänge zur Einschätzung von KI-Tutorensystemen

## 8.1 Zwischen Akzeptanz und Ablehnung – Hochschulen im KI-Dilemma

Muss Künstliche Intelligenz (KI) im Bildungswesen mehr Platz einnehmen und bleiben Potenziale bislang ungenutzt? Diese Frage lässt sich spätestens durch den Vormarsch von Künstlicher Intelligenz in nahezu alle Lebensbereiche kaum mehr unbeantwortet lassen.

### 8.1.1 Lernen neu gedacht – Wie KI den Bildungssektor verändert

Nachdem die KI mit der Verfügbarkeit von ChatGPT nochmal einen zusätzlichen Boost erhalten und im Alltag vieler Menschen spürbar Einzug gehalten hat, sind nun vermehrt auch Bildungseinrichtungen mit der Frage des Umgangs konfrontiert. So ist es wenig verwunderlich, dass auch das Bundesministerium für Bildung und Forschung zunächst im Jahre 2023 Fachforen eingerichtet, im Folgejahr eine Expertengruppe einberufen und im Jahr 2025 schließlich den Bildungsforschungstag der Frage „Wo stehen wir in zehn Jahren beim Thema KI in der Schule?“ gewidmet hat (Bundesministerium für Forschung und Bildung, 2025, online). Die bisherigen Erkenntnisse reichen für eine Handreichung zum Stand in Wissenschaft und Praxis (Scheiter et. al., 2025, online). Doch auch über die Schulbildung hinaus steht die Wissenschaft und insbesondere Hochschulen und Universitäten vor der Herausforderung, KI effizient und zielgerichtet einzusetzen. An diesem Punkt ist man auch bereits längst über die Debatte eines Verbots im wissenschaftlichen Kontext hinaus und diskutiert neben den Risiken auch gleichermaßen Chancen der KI-Nutzung. Hier steht insbesondere die Frage im Vordergrund, wie unter Zuhilfenahme der KI gelernt, geforscht und mit Erfolg studiert werden kann. In Ergänzung zu den zahlreichen herausgegebenen Richtlinien der Hochschulen zum geregelten Umgang mit KI (bspw. Gumprecht, S. & Meyer, S. & Peppel L., 2024, online), sollten die Chancen in Bezug auf wissenschaftliche Arbeit gleichermaßen Beachtung finden. In genau diesem Kontext bewegt sich die Forschung dieser Arbeit, mit der ein weiterer Beitrag zum Einsatz von KI an Hochschulen geleistet wird.

### **8.1.2 Studieren mit System – KI-Tutoren unter der Lupe**

Im Rahmen von Chancen der KI Nutzung an Hochschulen sind Lernplattformen ein relevanter Betrachtungsfokus. Neben den bereits gängigen Plattformen wie Moodle zum Austausch von Lernmaterialien zwischen Professoren und Studenten existieren weitere Online-Angebote der Hochschulen, um beispielsweise Prüfungen und Semesterbeiträge zu verwalten oder Formulare mit der Hochschule auszutauschen. Diese Plattformen können bereits als herkömmlich betrachtet werden und erhalten nun zunehmende Konkurrenz von KI gestützten Angeboten, die insbesondere darauf abzielen, dass Lern- und Studienerlebnis selbst zu revolutionieren. Die Hoffnungen liegen dabei auf einer effizienteren Wissensvermittlung, die durch den Einsatz von Künstlicher Intelligenz individualisiert und professionalisiert wird. Auch Prüfungsleistungen in Form von Wissenschaftlichen Arbeiten aber auch Klausuren sind zunehmend geprägt durch die Zuhilfenahme von KI-Lösungen. Beide Anwendungsbereiche bilden die Zielsetzung der Forschung und sollen dahingehend mit Hilfe quantitativer Methoden untersucht werden. Dabei spielt die Position der Studenten als Befragungsgruppe eine entscheidende Rolle, um Rückschlüsse auf die Akzeptanz und den weiteren Einzug KI-gestützter Lernmöglichkeiten im Hochschulkontext ziehen zu können.

### **8.1.3 Ersetzt KI den Dozenten? – Die zentrale Frage dieser Arbeit**

Zugrunde liegt dabei die Forschungsfrage, ob der Einsatz KI-gestützter Lernplattformen das Potenzial hat, klassische Lehrmethoden vollständig zu verdrängen. Dabei ist die Formulierung bewusst drastisch gewählt, um zu untersuchen, ob künftig nur noch KI Dozenten zur Vorbereitung und Ablage von Prüfungen im universitären Kontext zum Einsatz kommen. Zur Erörterung der Forschungsfrage ordnet die Arbeit zunächst den wissenschaftlichen Kontext ein, indem wichtige Definitionen beschrieben und existierende und artverwandte Studien dargestellt werden. In der Problemstellung werden anschließend die Forschungsfrage genauer beschrieben, die Hypothesen vorgestellt und die erwarteten Effekte dargestellt. Im darauffolgenden Kapitel steht die Methodik im Fokus der Betrachtung, indem das Design der Forschung aufgezeigt, die Stichprobe und Erhebungsinstrumente vorgestellt und schließlich die Durchführung der Erhebung betrachtet wird. Das anschließende Kapitel umfasst die Analyse der erhobenen Daten und leitet darauf basierend zur Interpretation der Ergebnisse über, bevor diese im letzten Kapitel abschließend diskutiert und kritisch eingordnet werden.

## **8.2 Wissen schafft Verständnis – Die Theorie hinter dem KI-Tutor**

### **8.2.1 Was genau ist ein KI-Tutor? Begriffe und Konzepte im Überblick**

Diese Kapitel sichert zunächst das Grundverständnis des Forschungskontextes, indem wichtige Begrifflichkeiten definiert und genauer erläutert werden.

#### **8.2.1.1 Der digitale Tutor – Lernen neu gedacht durch KI**

KI gestützt beschreibt die Vorgehensweisen, bei denen Künstliche Intelligenz zum Einsatz kommt, um den menschlichen Einsatz von Intelligenz zu unterstützen oder gar zu imitieren (Lackes, R. & Siepermann, M., o. J., online).

Übertragen auf Tutorenmodelle, welche ein Lernsystem mit individueller Anpassung an die Bedürfnisse sowie das Tempo des Lernenden im Kontext eines akademischen Studiums darstellen, bieten KI-gestützte Tutorenmodelle eine virtuelle Lernumgebung. Die Individualisierung der

Lerninhalte auf Basis der Lernaktivität des Studenten wird mit Hilfe der Künstlichen Intelligenz bereitgestellt und ständig aktuell gehalten. Die Lernumgebung ist zumeist auf einer Plattform eingebettet. Diese macht in der Regel nicht nur die relevanten Lerninhalte verfügbar, sondern stellt darüber hinaus die Aktualität und Individualität der Inhalte sowie sofortige Antworten und Erklärungen sicher und steht zudem rund um die Uhr zur Verfügung. Ziel ist es die individuellen Lernbedürfnisse der Nutzer möglichst genau abzudecken, um Lernziele hinsichtlich einer abzulegenden Prüfungsleistung mit maximalem Erfolg und Effizienz zu erreichen. Eine Erweiterung hierzu kann die Prüfungsleistung an sich sein, die ebenfalls in der virtuellen Umgebung durch die Interaktion mit der Künstlichen Intelligenz stattfindet (Schwaab, M. & Durian, A. (Hrsg.), 2025, S. 275).

#### **8.2.1.2     *Kognition trifft KI – Was effektives Lernen heute braucht***

Das wichtige und bereits beschriebene Ziel des Einsatzes von KI-gestützten Tutorensystemen ist die Effektivität beim Lernen zu fördern und den Lernprozess mit Individualisierung zu optimieren. Lerntheorie beschreibt dabei die Art und Weise des Wissenserwerbs von Individuen und soll darüber Aufschluss geben, wie Wissen durch Lernen angeeignet und schließlich in verschiedenen Situationen angewendet werden kann. Dabei wird Lerntheorie häufig in Kategorien eingeteilt. Behavioristisches Lernen wird als Zusammenspiel zwischen Reizen und Reaktionen gesehen und führt damit zu konditioniertem Verhalten auf Basis erlernter Erfahrung. Sofern der Wissenserwerb komplex und strukturiert geplant erfolgt, lässt sich von kognitiver Lerntheorie und den zugrunde liegenden Lernprozessen sprechen. Eine weitere Form ist das Soziale Lernen, dass durch die Interaktion in sozialen Konstrukten erfolgt. Darüber hinaus beschreibt die Organisationslerntheorie die Korrektur von Denkmustern und -modellen als wichtigen Teil des Lernprozesses (Bartscher, T. & Nissen, R., o. J., online). Grundlegend für den Untersuchungskontext dieser Arbeit wird die Kognitive Lerntheorie angesehen, da diese am ehesten den über einen KI-gestützten Tutor gesteuerten Lernprozess beschreibt. Sofern auf Basis dessen ein Lernerfolg einsetzt, lässt sich von Lerneffektivität sprechen. Indikatoren dafür sind insbesondere die langfristige Abrufbarkeit sowie Anwendbarkeit des erlernten Wissens und den Einsatz sowie die Transferleistung in anderen als der Lernsituation. Dabei sollte gleichermaßen die Motivation in der Lernsituation aufrechterhalten und der Zeitaufwand in angemessenem Rahmen gehalten werden. Diese Anforderung des effektiven Lernens gilt es gleichermaßen an KI-gestützte Tutorensysteme anzulegen. Dabei sind auch die Anwendung unterschiedlicher Lernmethoden sowie die Anpassung des Schwierigkeitsgrads von großer Bedeutung (Zendler, A., 2018, S. 159ff.).

#### **8.2.1.3     *Jeder lernt anders – Und KI passt sich an***

Auf Basis der Individualisierung und Anpassung innerhalb von Lernsystemen, um Lerneffektivität zu erreichen, spielen adaptive Lernsysteme eine entscheidende Rolle. Durch die Möglichkeit der kontinuierlichen Kontrolle und Bewertung des Lernerfolgs ist mit adaptiven Lernsystemen ein zentrales Element der Lerneffektivität erfüllt.

Als adaptives Lernsystem lässt sich ein Lernverfahren bezeichnen, das computergestützte Technologien verwendet, um einen auf die Bedürfnisse des Lernenden abgestimmten Lernpfad zu personalisieren. Auf Basis der im Lernfortschritt gesammelten Daten individualisieren Algorithmen die angebotenen Lerninhalte und stimmen diese auf den jeweiligen Lernfortschritt und das Tempo des Lernenden ab (Gligorea, I et. al., 2023, online). Der Einsatz von Künstlicher Intelligenz fördert dabei die Sicherstellung der Hauptkriterien von adaptiven Lernsystemen (Streicher, A., o. J., online). Zudem sind die laufende Aktualisierung und Optimierung der Inhalte sowie die Anwendung aktueller

Lernstandards und -inhalte wichtige Bestandteile von adaptiven Lernsystemen (Clarebout, G., Desmet, P. & Vandewaetere, M., 2010, online).

Im wachsenden Markt von KI-gestützten Lernplattformen unterscheiden sich diese zumeist durch die verschiedenen Schwerpunkte der Inhalte sowie die Features der Interaktion, welche die individuelle Lernerfahrung prägen (Kem, D., 2022, online). Für die vorliegende Arbeit werden Online Lernplattformen mit gängigen Individualisierungs- sowie Motivationsfeatures in die Betrachtung genommen. Unter diesen Aspekten soll der Einsatz von adaptiven Lernsystemen als KI-gestützte Tutorensysteme im Kontext eines akademischen Studiums im genaueren Fokus stehen.

## 8.2.2 Studienlage zu KI-Tutoren im Überblick

### 8.2.2.1 Wie KI-Tutoren den Lernerfolg steigern können

Eine Vielzahl von Studien setzt sich bereits mit der Frage auseinander, ob KI bzw. adaptive Lernsysteme Lernprozesse individualisieren und die Lernergebnisse optimieren können. Im Folgenden werden wesentliche Erkenntnisse aus aktueller Forschung sowie relevanten Studien präsentiert.

Die im Jahr 2024 von Sumei Hu durchgeführte Meta-Analyse basiert auf 36 experimentellen und quasi-experimentellen Einzelstudien, die in insgesamt 31 wissenschaftlichen Veröffentlichungen erschienen sind. Die Meta-Analyse ist ein Beitrag zu diesem Forschungsfeld und befasst sich mit dem KI-gestützten personalisierten Lernen bei Schülerinnen und Schülern. (Hu, S., 2024)

Ziel dieser Meta-Analyse war es herauszufinden, ob KI-gestütztes personalisiertes Lernen die Lernergebnisse im Vergleich zu traditionellen Lernmethoden signifikant verbessert. Zudem wurde untersucht, welchen Einfluss unterschiedliche EdTech-Anwendungen auf das personalisierte Lernen haben und inwieweit verschiedene Faktoren wie das Anwendungsszenario, die Anwendungsdauer, der Fachbereich und die Bildungsstufe die Auswirkungen von KI-gestütztem Lernen beeinflussen. (Hu, S., 2024)

Die Ergebnisse der Meta-Analyse zeigen, dass KI-gestütztes Lernen insgesamt einen moderaten positiven Effekt auf die Lernergebnisse hat. Die positiven Effekte zeigen sich vor allem in den Bereichen Wissenszuwachs, Kompetenzentwicklung und emotionaler Entwicklung, beispielsweise in Form von Motivation oder Lernfreude. Der Lernerfolg wird ebenfalls durch die Art der eingesetzten KI-basierten Lerntechnologien beeinflusst. Adaptive Systeme wirken sich stärker positiv aus als einfache standardisierte Anwendungen. Eine personalisierte, gezielt gestaltete Lernumgebung, die von einer KI begleitet wird, erzielt tendenziell bessere Lernergebnisse als andere Anwendungsszenarien wie das kollaborative Lernen in der Gruppe, das ebenfalls durch KI begleitet wird. (Hu, S., 2024)

Wie sich zeigt, können adaptive Systeme die Lernergebnisse von Lernenden positiv beeinflussen, sofern sie im richtigen Anwendungsszenario und über die entsprechende Anwendungsdauer eingesetzt werden. Doch wie sieht es bei Hochschulen aus?

Die im Jahr 2025 von Carlos Merino-Campos durchgeführte systematische Übersichtsarbeit basiert auf 45 Studien und leistet einen Beitrag zu diesem Forschungsfeld. Sie befasst sich mit KI-gestütztem, personalisiertem Lernen im Hochschulbereich und konzentriert sich auf die Förderung der Lernleistung sowie das Engagement der Studierenden (Merino-Campos, C. 2025).

Ziel war es, zu untersuchen, ob KI-gestütztes, personalisiertes Lernen Lernprozesse im Hochschulumfeld sinnvoll unterstützen kann. Dazu wurde der Einfluss adaptiver Systeme auf individuelle Lernpfade, Motivation, Lernerfolg und Selbststeuerung untersucht. (Merino-Campos, C. 2025)

Die Ergebnisse der systematischen Literaturübersicht zeigen, dass KI-gestütztes Lernen im Hochschulumfeld Potenzial hat. KI-Tutorensysteme, die adaptives Feedback nutzen, führen zu messbaren Verbesserungen der Lernergebnisse von Studierenden. Zudem bestätigen Studierende, dass sie durch die KI-Anwendungen, die in den Kursen verwendet wurden, motivierter sind und eine höhere Abschlussquote erreichen. (Merino-Campos, C. 2025)

Trotz der positiven Effekte auf die zuvor genannten Punkte kristallisieren sich bei der Verwendung derartiger adaptiver Systeme Herausforderungen heraus. Dazu zählen die Themen Datenschutz, algorithmische Verzerrungen sowie unzureichende Lehrkräfte und eine mangelhafte Infrastruktur. Vor allem im Hochschulumfeld verschärfen höhere Datenmengen die Problematik des Datenschutzes. Es muss Vertrauen geschaffen und Lehrkräfte müssen gezielt geschult werden. (Merino-Campos, C. 2025)

### **8.2.2.2 Nützlichkeit von KI-Tutoren für Studierende**

Die oben dargestellten Studien zeigen, dass KI-Tutoren die Lernergebnisse sowie Motivation steigern können. Doch wie sieht es mit der Akzeptanz von KI im Hochschulkontext aus?

Die explorative Studie von Kieslich, Marcinkowski und Starke (2019), die sich mit der Akzeptanz und Nutzung von KI-Anwendungen befasst, liefert in diesem Umfeld zentrale Erkenntnisse. Dafür wurden 304 Studierende der Heinrich-Heine-Universität Düsseldorf auf Basis einer standardisierten Erhebung befragt. Die Akzeptanz von KI ist stark anwendungsabhängig. (Kieslich, K. et al., 2019)

Dabei werden KI-Anwendungen, die Vorhersagen treffen, verschiedene Analysen durchführen oder Empfehlungen abgeben können, positiver bewertet als Anwendungen, die selbstständig Entscheidungen treffen können. (Kieslich, K. et al., 2019)

Zudem fordern die Studierenden die Modernisierung der Technik in den Räumlichkeiten sowie eine Erweiterung des Lehrangebots. Gerade im Bereich der Learning Analytics und KI-Tutoren versprechen sich die Studienteilnehmer einen persönlichen Nutzen. Learning Analytics sind Anwendungen, die detaillierte Informationen über das Lern- und Studierverhalten analysieren und auf dieser Grundlage gezielte Maßnahmen ableiten können. (Kieslich, K. et al., 2019)

### **8.2.2.3 Lernen ohne Budget**

Die oben aufgeführten Studien zeigen, dass adaptive Systeme die Lernergebnisse verbessern können, bei den Studierenden Akzeptanz finden und aus deren Perspektive als nützlich angesehen werden. In diesem Kontext stellt sich die Frage, ob Studierende bereit wären, für KI-Anwendungen einen monetären Beitrag zu zahlen.

Genau mit dieser Frage setzt sich die Studie „Students' Willingness to Pay for Access to ChatGPT“ von Lupa-Wójcik, I. (2024) auseinander (Lupa-Wójcik ,2024). Dafür wurden 424 Studierende der Pädagogischen Universität Krakau befragt, ob sie bereit wären, für den Zugang zu ChatGPT zu zahlen.

Mehr als ein Drittel der Studierenden lehnte Gebühren ab oder akzeptierte nur sehr niedrige Preise (10 PLN = 2,35 €). Ein realistischer Preis scheint unter 3 € pro Monat zu liegen. Am stärksten beeinflussen die Variablen Geschlecht, kommerzielle Anwendung von KI sowie finanzielle Variablen die Zahlungsbereitschaft. (Lupa-Wójcik ,2024)

## **8.3 KI statt Präsenz?**

Die relevanten Fragen, die sich im Rahmen dieser Forschung nun stellen, sind, ob Studierende solche Systeme im Kontext der Prüfungsvorbereitung, der Prüfungsformate und der Leistungsbewertung wirklich akzeptieren und aktiv nutzen würden.

Die vorliegende Arbeit beschäftigt sich damit, inwieweit KI-Tutoren traditionelle Präsenzveranstaltungen und Prüfungsformate an Hochschulen ersetzen können. Da diese Frage viele Dimensionen umfasst, konzentriert sich die Analyse auf die Akzeptanz, Nutzung, Bewertung und Zahlungsbereitschaft sowie zeitliche Aspekte für einen möglichen Ersatz klassischer Lehrmethoden.

Die zentrale Forschungsfrage ist: Kommen künftig nur noch KI Dozenten zur Vorbereitung und Ablage von Prüfungen im universitären Kontext zum Einsatz?

Um diese Frage zu beantworten, werden verschiedene Aspekte des studentischen Verhaltens und der Einstellungen gegenüber KI-Tutoren untersucht. Daraus ergeben sich folgende untergeordnete Forschungsfragen.

### **8.3.1 Forschungsfragen**

1. Wie hoch ist die Akzeptanz von KI-Tutoren zur Prüfungsvorbereitung?
2. Welchen Mehrwert sehen Studierende in KI-Tutoren im Vergleich zu klassischen Lernmitteln?
3. Wie hoch ist die Bereitschaft der Studierenden, Prüfungen in einem KI-gestützten, adaptiven Format ohne menschliche Interaktion zu absolvieren?
4. Würden Studierende akzeptieren, dass eine KI ihre Prüfungsleistungen bewertet und benotet?
5. Wie viel Zeit sind Studierende bereit, wöchentlich mit einem KI-Tutor zu verbringen?
6. Welche Zahlungsbereitschaft besteht für zusätzliche KI-basierte Lernfunktionen wie individuelles Feedback oder Simulationen?
7. Besteht ein Zusammenhang zwischen der Akzeptanz von KI-Tutoren für Prüfungen und der Zahlungsbereitschaft?

### **8.3.2 Hypothesen**

H1: Studierende zeigen eine positive Nutzungsbereitschaft ( $\geq 4$  auf einer 5-stufigen Likert-Skala) für KI-Tutoren als Werkzeug in der Prüfungsvorbereitung.

H2: Studierende, die einen höheren Mehrwert in KI-Tutoren gegenüber klassischen Lehrmitteln erkennen, zeigen eine höhere Nutzungsbereitschaft.

H3: Es besteht ein positiver Zusammenhang zwischen der Nutzungsbereitschaft von KI-gestützten Tutorensystemen und der Zustimmung zur Bewertung durch KI.

H4: Es besteht ein positiver Zusammenhang zwischen der Akzeptanz von KI-gestützten Prüfungsformaten und der Zustimmung zur Bewertung durch KI.

H5: Studierende, die KI-gestützte Prüfungsformate akzeptieren, zeigen eine höhere Zahlungsbereitschaft für zusätzliche Funktionen.

H6: Es besteht ein positiver Zusammenhang zwischen der angegebenen Lernzeit mit dem KI-Tutor und der Akzeptanz von KI-Nutzung im Bildungskontext.

### **8.3.3 Erwartete Effekte**

Grundsätzlich wird auf Basis aktueller Studien davon ausgegangen, dass Studierende dem Einsatz von KI zur Optimierung ihrer Lernprozesse offen gegenüberstehen, insbesondere dann, wenn ein individueller Nutzen entsteht. Die Analysen von HU (2024) und Merino-Campus zeigen positive Aspekte bei der Verwendung adaptiver KI-Systeme. Diese Systeme haben vor allem positive Effekte auf Motivation und Lernerfolg. Aufgrund der übereinstimmenden Ergebnisse wird davon ausgegangen, dass Studierende eine hohe Bereitschaft zeigen, KI-Tutoren zur Prüfungsvorbereitung zu nutzen. Ein weiterer erwarteter Effekt ist, dass Studierende, die eine Nutzungsbereitschaft aufweisen auch offen für eine Bewertung durch KI-gestützte Systeme sind. Hingegen ist davon auszugehen das Studierende, die KI-gestützte Prüfungsformate akzeptieren auch die Bewertung der Ergebnisse der KI überlassen. Aufgrund der oben aufgeführten Studie von Lupa-Wójcik (2024) wird auch davon ausgegangen, dass Studierende eher nicht bereit sind, für KI-Tutoren zu zahlen, allenfalls einen geringen Betrag.

## **8.4 Der Weg zur Antwort**

Nachdem die Problemstellung der Erhebung an dieser Stelle bekannt ist, soll dieses Kapitel herausarbeiten, welcher methodische Ansatz verwendet wird, um die Forschung durchzuführen.

### **8.4.1 Forschungsdesign**

Mit der Forschungsfrage, ob KI-gestützte Tutorensysteme klassische Lehrmethoden im universitären Kontext gänzlich ersetzen können, sollen diverse kausale Zusammenhänge gemessen werden. Dafür wird ein quantitatives Forschungsdesign gewählt. Auf diese Weise können die Befragungsergebnisse genutzt werden, um Zusammenhänge zu analysieren und damit Rückschlüsse auf eine größere Allgemeinheit im Sinne der Forschungsfrage zu ziehen. Vorteil des quantitativen Designs ist, dass die Verdichtung der Forschungsfrage mittels der aufgestellten Hypothesen anhand von Zahlen analysiert werden kann. So kann der recht komplexe Zusammenhang zwischen KI-gestützten Tuto ren und klassischen Lehrveranstaltungen unter Reduktion der Komplexität besser verstanden werden.

Die gewählte Methode für die quantitative Erhebung ist eine standardisierte Befragung, die mittels hypothesenprüfenden Verfahrens analysiert wird (Fahrmeir et. al., 2023, S. 12ff.). Hierfür kommt dieselbe Online-Umfrage mit standardisiertem Fragebogen, wie in den Kapiteln zuvor beschrieben, zum Einsatz. Diese Methodik eignet sich insbesondere, um eine möglichst große Grundgesamtheit zu erreichen und dabei die spezifischen Positionen der Stichprobe auf die Prüfung der Forschungsfrage anzuwenden. Zu den Details der weiteren methodischen Aspekte ist das allgemeine Kapitel zur Methodik heranzuziehen.

Für die Datenanalyse wird ein zweistufiger Ansatz gewählt. Zunächst werden mittels deskriptiver Statistik die Häufigkeitsverteilungen der Antworten analysiert und durch geeignete Visualisierungen grafisch aufbereitet. In einem zweiten Schritt erfolgt die statistische Auswertung durch Hypothesentests. Hierfür wird die Spearman-Korrelation verwendet, die besonders geeignet ist, da sie als nicht-parametrisches Verfahren keine Normalverteilung der Daten voraussetzt und sich gut für die Analyse von Likert-Skalen eignet (Fahrmeir et. al., 2023, S. 422f.). Die Beurteilung der Hypothesen erfolgt durch die Analyse der Korrelationskoeffizienten in Verbindung mit den entsprechenden p-Werten. Dabei wird ein Signifikanzniveau von  $\alpha = 0,05$  als Schwellenwert für die statistische Signifikanz festgelegt. P-Werte unterhalb dieses Schwellenwerts führen zur

Beibehaltung der jeweiligen Hypothese, während höhere p-Werte zu deren Verwerfung führen. Für die Analyse der metrischen Daten, also der zeit- und kostenbezogenen Angaben, werden zusätzlich Durchschnittswerte berechnet und Verteilungsanalysen durchgeführt. Die gesamte statistische Auswertung wird mittels Python durchgeführt, wobei insbesondere die Bibliotheken pandas für die Datenverarbeitung, scipy für statistische Berechnungen und matplotlib sowie seaborn für die grafische Darstellung zum Einsatz kommen (pandas, o. J., online & seaborn, o. J., online & scipy, o. J., online & matplotlib, o. J., online).

Um die wissenschaftliche Qualität der quantitativen Erhebung sicherzustellen, werden die zentralen Gütekriterien quantitativer Forschung zugrunde gelegt.

#### 8.4.2 Stichprobe

Die Beschreibung der Stichprobe ist bereits im allgemeinen Kapitel zur Methodik näher betrachtet und entfällt an dieser Stelle.

#### 8.4.3 Erhebungsinstrumente

Für die allgemeine Erhebung sind folgende Fragen aus diesem Kapitel eingeflossen. Die theoretischen Annahmen und Konstrukte werden im Rahmen der quantitativen Forschung systematisch in messbare Variablen überführt. Als Basis dienen die im vorherigen Kapitel definierten Hypothesen. Bei der Operationalisierung sind die eindeutige Messbarkeit und genaue Abbildung der theoretischen Annahmen sowie die statistische Auswertbarkeit im Vordergrund. Die Fragen sind den Hypothesen folgendermaßen zugeordnet (Lange, T. & Mosler, K., 2017, S.64):

Frage im Online-Fragebogen	Hypothesen	Zielvariable(n)
1. Ich bin bereit, mit einem KI-Tutor meine Prüfungsvorbereitung zu optimieren.	H1, H2, H3, H6	Zustimmung zur KI-Nutzung (KI_Nutzung)
2. Ich sehe einen deutlichen Mehrwert in der Integration eines KI-Tutors im Vergleich zu klassischen Lernmitteln und Lehrveranstaltungen und kann mir eine Prüfungsvorbereitung nur (ohne Skripte, Bücher, Vorlesungen) damit vorstellen.	H2, H3, H4, H6	Mehrwertwahrnehmung & Nutzungsbereitschaft (KI_Mehrwert & KI_Nutzung)
3. Ich bin bereit, eine Prüfungsleistung in Form eines KI-gestützten, interaktiven Abfrageformats zu absolvieren – z. B. im Dialog mit einem adaptiven KI-Tutor – anstelle einer klassischen Klausur oder Hausarbeit.	H4, H5	Akzeptanz KI-Prüfungsformat (KI_Pruefung)
4. Ich kann mir vorstellen, dass ein KI-Tutor meine Prüfungsleistungen bewertet und eine Note vergibt.	H3, H4	Zustimmung zu KI-Bewertung (KI_Bewertung)
5. Wie viel Zeit pro Woche und Modul sind Sie bereit mit dem KI-Tutor zu interagieren (systemseitiges Tracking und Kontrolle durch Hochschule), bevor Sie zu einer Prüfung zugelassen werden?	H6	Lernzeit mit KI-Tutor in Stunden (KI_Zeit)
6. Ich wäre bereit, für zusätzliche Tutor-Funktionen wie Simulationen, individuelles Feedback oder adaptive Lernpläne einen monetären Beitrag pro Monat zu leisten.	H5	Zahlungsbereitschaft (KI_Zahlung)

Tabelle 8.1: Operationalisierung Hypothesen und Fragen und Zielvariablen

## 8.5 Nutzung ja – Zahlung nein

Anhand der beschriebenen Vorgehensweise zur Datenverarbeitung beschäftigt sich dieses Kapitel mit der statistischen Vorgehensweise und bereitet die Daten mit den relevanten Instrumenten auf. Darauf basierend erfolgt eine kritische Betrachtung der Hypothesen und die Interpretation der Ergebnisse auf Basis der Forschungsfrage.

### 8.5.1 Datenanalyse

Die Analyse der Stichprobe ( $n=66$ ) zeigt eine grundsätzlich positive Einstellung zur KI-Nutzung im Hochschulkontext. Die demografische Analyse zeigt eine mehrheitlich männliche Teilnehmerschaft, eine Altersverteilung um 22 Jahre ( $M=22,4$ ) und ein durchschnittliches Semester von 3,9, was auf erfahrene Bachelorstudierende hindeutet.

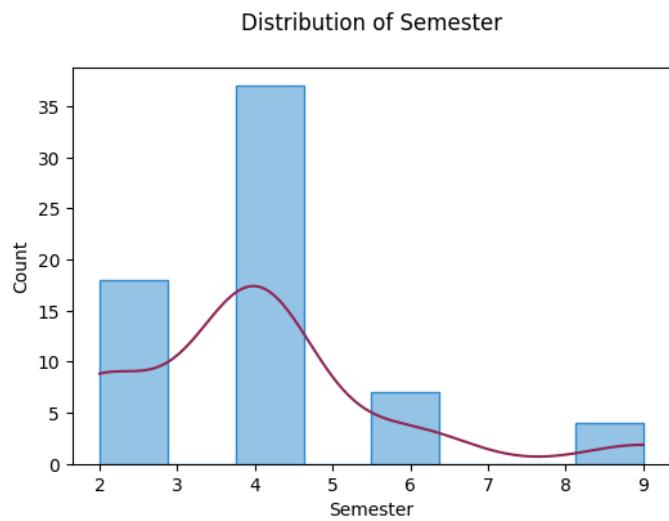


Abbildung 8.2: Häufigkeitsverteilung Semester

Die allgemeine KI-Erfahrung der Teilnehmer unabhängig von KI-gestützten Tutorensystemen ist hoch (>80% im oberen Spektrum der 7er-Skala), mit entsprechend häufiger Nutzung (80% "Häufig" oder "Bei jeder wissenschaftlichen Arbeit"), was die Stichprobe für die Forschungsfrage qualifiziert.

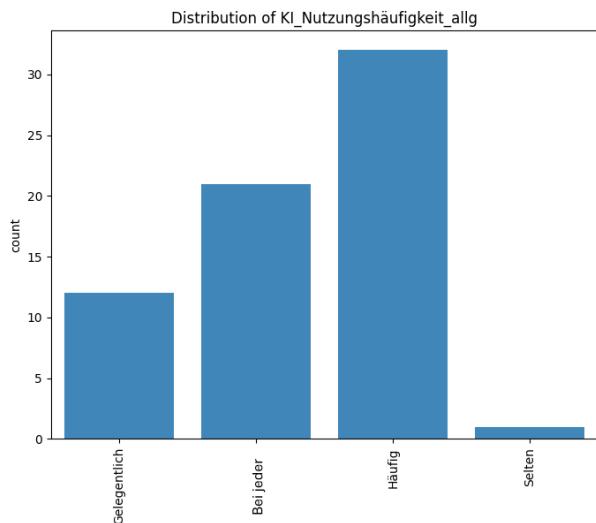


Abbildung 8.3: Häufigkeitsverteilung der Antworten zur KI-Nutzungshäufigkeit

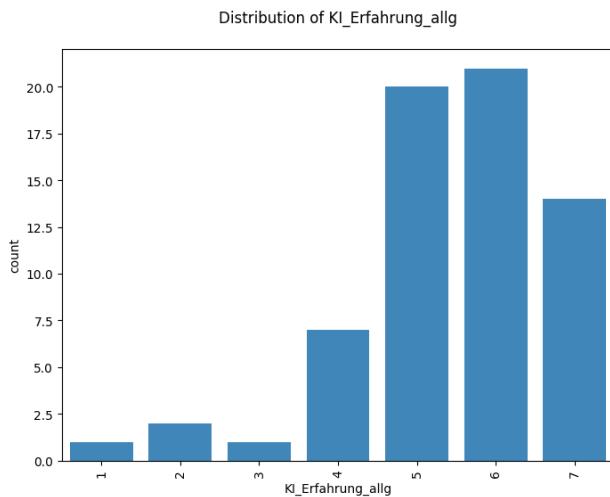


Abbildung 8.4: Häufigkeitsverteilung der Antworten zu KI-Erfahrung

Um die Analyse der Daten im Kontext der Forschungsfrage bzgl. KI-gestützter Tutorensysteme vorzubereiten, werden die Voraussetzungen und die Datenqualität zunächst genauer beleuchtet. Die Normalverteilungstests zeigen, dass nur KI\_Mehrwert ( $p=0,0769$ ) und KI\_Pruefung ( $p=0,1595$ ) normalverteilt sind, während andere Variablen signifikante Abweichungen aufweisen. Die deskriptiven Statistiken zeigen eine positive Einstellung zur KI-Nutzung ( $M=3,95$ ,  $SD=1,14$ ), moderate Mehrwerteinschätzung ( $M=3,17$ ,  $SD=1,26$ ) und neutrale Prüfungsbewertung ( $M=3,06$ ,  $SD=1,26$ ).

Die Spearman-Korrelationsanalyse (Fahrmeir et. al., 2023, S. 142ff.) zeigt signifikante positive Zusammenhänge, insbesondere zwischen KI\_Mehrwert und KI\_Pruefung ( $rs=0,567$ ), KI\_Bewertung und KI\_Pruefung ( $rs=0,472$ ) sowie KI\_Nutzung und KI\_Mehrwert ( $rs=0,432$ ). Bemerkenswert ist, dass die numerischen Variablen KI\_Zeit\_num und KI\_Zahlung\_num keine signifikanten Korrelationen ( $|rs| < 0,3$ ) mit den anderen Variablen aufweisen.

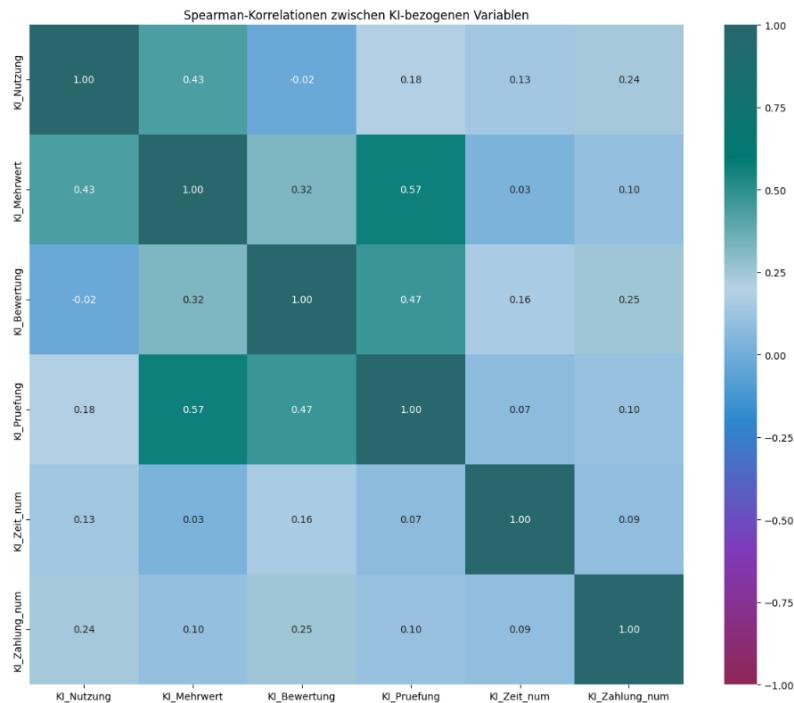


Abbildung 8.5 Spearman-Korrelationsanalyse der Variablen im Kontext KI-gestützte Tutorensysteme

Für die weitere Analyse werden nicht-parametrische Verfahren verwendet, indem die Korrelation der Hypothesen-Variablen nach Spearman analysiert wird. Anhand dieses Verfahrens kann über die Annahme bzw. Ablehnung der Hypothesen entschieden werden. Es wird daher davon ausgegangen, dass die 5er Likert-Skalen ordinal sind, also die Abstände zwischen den Antwortoptionen nicht als gleich groß angesehen werden.

### 8.5.2 Statistische Tests

Zunächst wird der Binomial-/Chi-Quadrat-Test zur Prüfung der Hypothese 1 angewendet (Lange, T. & Mosler, K., 2017, S.71). Die Analyse der Variable KI-Nutzung zeigt eine signifikant positive Einstellung der Befragten. Die Verteilung der Antworten auf der 5-stufigen Likert-Skala weist eine deutliche Tendenz zur Zustimmung auf, wobei 68,2% der Teilnehmer eine positive Haltung (Werte  $\geq 4$ ) ausdrücken. Der größte Anteil entfällt dabei auf die höchste Zustimmungskategorie (42,4%, Teil der Stichprobe ( $n = 28$ ), gefolgt von "Stimme eher zu" (25,8%,  $n=17$ ). Ein Binomialtest bestätigt, dass diese Verteilung signifikant von einer zufälligen 50:50-Verteilung abweicht ( $p=0,0043$ ). Diese Beobachtung wird durch einen Chi-Quadrat-Test unterstützt ( $\chi^2=8,7273$ ,  $p=0,0031$ ), der ebenfalls eine signifikante Abweichung von einer Gleichverteilung nachweist. Neutrale Positionen nehmen 21,2% ( $n=14$ ) der Befragten ein, während nur eine Minderheit von 10,6% ( $n=7$ ) eine ablehnende Haltung zeigt. Diese rechtschiefe Verteilung deutet auf eine überwiegend positive Einstellung zur KI-Nutzung in der untersuchten Stichprobe hin, wobei die statistische Signifikanz beider Tests ( $p < 0,05$ ) eine robuste Schlussfolgerung zulässt. Die Hypothese 1, dass Studierende eine positive Nutzungsbereitschaft für KI-gestützte Tutorensysteme als Werkzeug in der Prüfungsvorbereitung zeigen, wird damit bestätigt.

Im Folgenden werden die Hypothesen 2 bis 6 mit Hilfe der Spearman Korrelation analysiert. Die Hypothese 2 zeigt dabei, dass diese auf Basis der erhobenen Daten nicht angenommen werden kann.

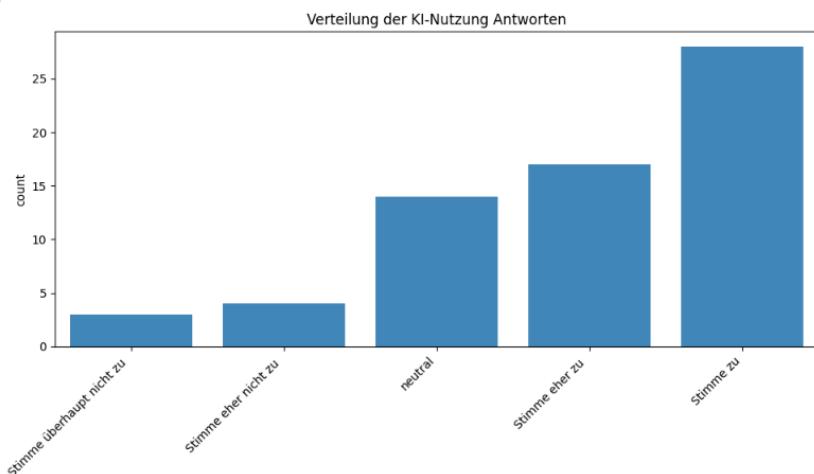


Abbildung 8.6 Häufigkeitsverteilung der Antworten zur Variable KI-Nutzung

Die vermutete positive Korrelation zwischen wahrgenommenem Mehrwert und Nutzungsbereitschaft konnte nicht bestätigt werden ( $rs = 0.0409$ ,  $p = 0.7445 > 0.05$ ). Auch Hypothese 3 führt nach dem Test der Spearman Korrelation zur Ablehnung, da keine signifikante Korrelation zwischen Nutzungsbereitschaft und Bewertung festgestellt ( $rs = -0.0426$ ,  $p = 0.7342 > 0.05$ ) werden kann. Die schwach negative Korrelation ist nicht aussagekräftig und deutet auf die Unabhängigkeit dieser beiden Dimensionen hin. Darauf basierend erweitert Hypothese 4 die Prüfung hinsichtlich der Akzeptanz von Prüfung und Bewertung durch KI-gestützte Tutorensysteme. Hier zeigt sich ein signifikant positiver Zusammenhang ( $rs = 0.3269$ ,  $p = 0.0074 < 0.05$ ).

	Hypothese	Testverfahren	Statistik	p-Wert	Ergebnis
0	H1	Binomial/Chi <sup>2</sup>	8.727300	0.003100	Angenommen
1	H2	Spearman	0.040900	0.744500	Abgelehnt
2	H3	Spearman	-0.042600	0.734200	Abgelehnt
3	H4	Spearman	0.326900	0.007400	Angenommen
4	H5	Spearman	-0.229200	0.064100	Abgelehnt
5	H6	Spearman	0.046500	0.710800	Abgelehnt

Abbildung 8.7 Zusammengefasste Darstellung des Hypothesentests auf Basis der Spearman-Korrelationsanalyse

Die Korrelationsanalyse zu den beiden numerischen Variablen zeigt keine signifikanten Zusammenhänge. Die Analyse der Hypothese 5 zeigt einen schwach negativen, jedoch nicht signifikanten Zusammenhang ( $rs = -0.2292$ ,  $p = 0.0641 > 0.05$ ). Auch die Analyse des zeitlichen

Aspektes (Zeit-Nutzungs-Beziehung  $rs = 0.0465$ ,  $p = 0.7108 > 0.05$ ) weist darauf hin, dass die investierte Zeit keinen nachweisbaren Einfluss auf die Nutzungsbereitschaft hat. Damit lässt sich zusammenfassen, dass zwei der sechs aufgestellten Hypothesen angenommen werden können. Neben der positiven Einstellung zur KI-Nutzung herrscht ein Zusammenhang zwischen Prüfungsakzeptanz und KI-gestützter Bewertung. Besonders bemerkenswert ist die tendenzielle negative Beziehung zwischen Prüfungsakzeptanz und Zahlungsbereitschaft (Hypothese 5).

Die Kruskal-Wallis-Tests (Fahrmeir et. al., 2023, S. 536) zeigen hochsignifikante Gruppenunterschiede ( $p < 0,001$ ,  $\eta^2 = 1,000$ ) mit konsistenter Differenzierung (mittlere Ränge 33,50,  $SD \approx 18$ ). Die Analyse der KI-Nutzungsbereitschaft zeigt eine deutlich positive Tendenz, wobei die größte Gruppe ( $n = 28$ ) maximale Zustimmung aufweist und ein systematischer Anstieg der mittleren Ränge von 2.00 bis 14.50 zu beobachten ist. Bei der Bewertung des KI-Mehrwerts und der KI-gestützten Prüfungsformate zeigt sich eine symmetrischere Verteilung mit Schwerpunkt im neutralen Bereich (jeweils  $n = 24$  bzw.  $n = 25$  bei Median = 3.00), während die allgemeine KI-Bewertung eine polarisierte Struktur mit Häufungen bei den Extremwerten aufweist (jeweils  $n = 19$  bei minimaler und mittlerer Ausprägung). Die durchgängig signifikanten paarweisen Vergleiche mit sehr niedrigen p-Werten (alle  $p < 0.001$  nach Bonferroni-Korrektur) bestätigen die Robustheit der gefundenen Gruppenunterschiede. Diese Ergebnisse unterstreichen die heterogene, aber strukturierte Wahrnehmung von KI-Technologien im Hochschulkontext, wobei insbesondere die hohe Nutzungsbereitschaft bei gleichzeitig differenzierter Bewertung der praktischen Implementierungsaspekte heraussticht. Die Ergebnisse deuten auf ein komplexes Akzeptanzmuster hin, das sowohl Offenheit für KI-Technologien als auch differenzierte Bewertung praktischer Aspekte umfasst.

### 8.5.3 Ergebnisinterpretation

Die Auswertung der gesammelten Daten ergibt ein vielseitiges Bild hinsichtlich der Nutzung und Akzeptanz von KI-gestützten Tutoren-Systemen im Hochschulbereich. Insgesamt kann festgehalten werden, dass die analysierten Stichproben eine positive Einstellung zum Einsatz KI-gestützter Tutoren in der Prüfungsvorbereitung aufweisen. Hypothese 1 wird hier nicht abgelehnt, was im Einklang mit den Ergebnissen der explorativen Studie von Kieslich et al. (2019) steht. In dieser Studie äußerten sich Studierende annehmend gegenüber KI-Anwendungen bzw. KI-Tutoren, wenn diese Systeme das Ziel verfolgten, zu analysieren, vorherzusagen oder zu empfehlen.

Hypothese 2 zufolge ist die Wahrnehmung eines Mehrwerts von KI-Tutoren nicht ausreichend, um eine höhere Bereitschaft zur Nutzung durch KI-Tutoren zu gewährleisten. Daraus lässt sich schließen, dass die Nutzungsbereitschaft offenbar weniger vom wahrgenommenen Mehrwert abhängt und stattdessen durch andere Faktoren beeinflusst wird. Die Studierenden sehen zwar einen potenziellen Nutzen in KI-Tutoren, aber das führt nicht automatisch zur aktiven Nutzung. Weiterhin könnte eine hohe Wertschätzung von klassischen Lehrmitteln, wie Vorlesungen, Skripte und Bücher der Grund der Ablehnung der Hypothese sein. Dies deutet daraufhin, dass KI-Tutoren nicht als vollständiger Ersatz für Studenten, sondern mehr als ergänzendes Lehrmittel wahrgenommen werden.

Die Ergebnisse für die Hypothese 3 deuten darauf hin, dass eine Bereitschaft zur Nutzung von KI-Tutoren nicht einhergeht mit der Bewertung durch KI-Systeme. Trotz der Offenheit der Studierenden zur Nutzung von KI-Tutoren, sind die Studierenden offenbar zurückhaltend, wenn es um die Beurteilung ihrer Leistungen geht. Dies ist ebenfalls im Einklang mit der Studie Kieslich et al. (2019), waren gegenüber KI die Entscheidungen negativer eingestellt, als KI die nur analysiert und Empfehlungen gibt. Dies könnte auf vorhandene Fehleranfälligkeit, Bias oder mangelnde

Transparenz zurückzuführen sein. KI-Tutoren werden also eher als Tool akzeptiert, jedoch weniger als bewertende Instanz in Prüfungssituationen.

Hypothese 4, die prüft, ob Studierende, die eine höhere Akzeptanz gegenüber KI-gestützten Prüfungsformaten haben, auch offen für die Bewertung durch diese sind, wurde bestätigt. Das bedeutet, dass die Akzeptanz von KI-gestützten Prüfungsformaten mit einer Akzeptanz der Bewertung durch KI einhergeht.

Wie auch in der Studie von Lupa-Wójcik (2024) festgestellt wurde, bleibt die Zahlungsbereitschaft ein kritischer Faktor. Ein Drittel (35 %) der durch Lupa-Wójcik Befragten ist nicht bereit, für KI-Anwendungen bzw. ChatGPT einen monetären Beitrag zu leisten. In der vorliegenden Analyse ist es ähnlich: 24 der 66 Stichprobenteilnehmer, also 36 %, waren nicht bereit, Geld für einen KI-Tutor auszugeben. Bei den Studierenden, die Zahlungsbereitschaft zeigen, liegt der durchschnittliche Betrag bei 5,10 €, was über dem Mittelwert der Vergleichsstudie liegt. Zu beachten ist jedoch, dass sich die Abfrageformate und Preisspannen zwischen den beiden Studien unterscheiden. In der folgenden Arbeit wurde eine höhere Preisspanne bis über 20 € verwendet, was einen direkten Vergleich erschwert. Der Mittelwert von 5,10 € ist im Vergleich zur gesetzten Preisspanne trotzdem als gering anzusehen. H5 wurde somit auch abgelehnt, was darauf hindeutet, dass Studenten mit höherer Prüfungsakzeptanz tendenziell eher nicht bereit sind, für KI-Tutoren Geld auszugeben. Die negative Tendenz deutet darauf hin, dass eine höhere Prüfungsakzeptanz tendenziell mit geringerer Zahlungsbereitschaft einhergeht.

Zudem wurde H6 abgelehnt. Dies deutet darauf hin, dass eine positive Haltung gegenüber der KI-Nutzung nicht automatisch eine längere Nutzung bedeutet. Der Mittelwert von 3,24 zeigt, dass die Studenten grundsätzlich bereit sind, Zeit pro Woche zu investieren, was jedoch nicht direkt mit der Zustimmung zur KI-Nutzung zusammenhängt. Eine Standardabweichung von 2,53 zeigt, dass es große Unterschiede im Nutzungsverhalten der Studierenden gibt.

Dies deutet darauf hin, dass die Einstellung zur KI-Nutzung und deren Bewertung weitgehend unabhängig von zeitlichen und monetären Aspekten ist. Die positiven Korrelationen zwischen den EinstellungsvARIABLEN legen nahe, dass eine positive Haltung in einem Aspekt der KI-Nutzung tendenziell mit positiven Einschätzungen in anderen Bereichen einhergeht.

## 8.6 KI überzeugt – mit Grenzen

### 8.6.1 Kernerkenntnisse

Auf Basis der Ergebnisse der vorliegenden Analyse kann geschlossen werden, dass Studierende grundsätzlich eine aufgeschlossene Haltung gegenüber KI-Tutoren im Hochschulumfeld haben, insbesondere im Zusammenhang mit der Prüfungsvorbereitung. Die grundsätzliche Bereitschaft KI-Tutoren zu nutzen, wird statistisch signifikant mit der Hypothese 1 bestätigt. Dies macht deutlich, dass die Studierenden das Potenzial solcher Systeme, vor allem in Bezug auf zeitliche Flexibilität, individuelles und effizientes Lernen, erkennen.

Zugleich die Analyse, dass ein empfundener Mehrwert nicht unbedingt mit einer aktiven Verwendung von KI-Tutoren korreliert (Hypothese 2 abgelehnt). KI-Tutoren werden in der Studie eher als ergänzendes Lernangebot angesehen und nicht als Ersatz für klassische Lern- und Lehrmethoden. Des Weiteren zeigt sich, dass eine vorhandene Nutzungsbereitschaft nicht mit einer Bewertung durch eine KI einhergeht (Hypothese 3 abgelehnt). Hingegen wurde eine positive Korrelation zwischen der Akzeptanz von KI-gestützten Prüfungsformaten sowie der Akzeptanz von KI-basierter Bewertung bestätigt (Hypothese 4).

Die Zahlungsbereitschaft der Studierenden bleibt ebenfalls begrenzt und nur zwei Drittel der Studierenden äußern sich überhaupt bereit, einen geringen monatlichen Betrag ( $M=5,10$ ) für einen KI-Tutor zu leisten (Hypothese 5 abgelehnt). Die mögliche Lernzeit, die ein Student für die Prüfung mit einem KI-Tutor aufbringt, steht dabei nicht im direkten Zusammenhang mit der Nutzungsbereitschaft von KI-Tutoren (Hypothese 6 abgelehnt).

### 8.6.2 Handlungsempfehlungen

Auf Grundlage der Ergebnisse der Analyse lassen sich folgende Handlungsempfehlungen ableiten:

Einbinden statt ersetzen: Bestehende Lernmethoden sollen durch KI-Tutoren als ergänzendes Werkzeug in gemischte Lernsettings eingebettet werden. Dieser Ansatz könnte die Akzeptanz bei Studenten steigern.

Rabatt-Angebote: Studenten sollten vergünstigte Angebote oder zunächst kostenfreie Versionen eines KI-Tutors angeboten werden. Aufgrund der geringen Zahlungsbereitschaft von Studierenden können Einstiegshürden in Form von Kosten die Nutzung und Akzeptanz verhindern.

Schulungen und Aufklärungen: Die Studierenden sollten über die Funktionsweise, Chancen und auch Grenzen von KI-Tutoren bzw. KI-gestützten Systemen aufgeklärt werden. Schulungen der Studenten, aber auch der Dozenten könnte eine solche Aufklärung gewährleisten.

### 8.6.3 Ausblick

Grundsätzlich zeigen die Ergebnisse, dass KI-Tutoren an Hochschulen akzeptiert werden, ein vollständiges Vertrauen und eine voluminöse Nutzung liegen jedoch nicht vor. Zukünftig ist es wichtig, wie Hochschulen die Entwicklung der KI-Anwendung begleiten. Für zukünftige Forschungen könnten folgende Gesichtspunkte vertieft werden:

1. Wie sieht die Akzeptanz von KI-Tutoren in verschiedenen Studiengängen aus?
2. Welche Lernergebnisse erzielt man mit einem KI-Tutoren über einen längeren Zeitraum?

Zusammenfassend kann festgestellt werden, dass KI-Tutoren das Potenzial besitzen, Lernprozesse im Hochschulkontext grundlegend zu verändern – jedoch nur, wenn ihre Einführung vorsichtig, differenziert und nutzerorientiert gestaltet wird.

## 8.7 Literaturverzeichnis

Bartscher, T. & Nissen, R. (o. J.). Lerntheorie. <https://wirtschaftslexikon.gabler.de/definition/lerntheorien-37820/version-261250>, zugegriffen am 20.05.2025.

Bundesministerium für Forschung und Bildung (2025). Begleitprozess „Künstliche Intelligenz im Bildungsbereich“. <https://www.empirische-bildungsforschung-bmbf.de/de/Begleitprozess-Kunstliche-Intelligenz-im-Bildungsbereich-2274.html>, zugegriffen am 14.06.2025.

Clarebout, G., Desmet, P. & Vandewaetere, M. (2010). The contribution of learner characteristics in the development of computer-based adaptive learning environments. <https://www.sciencedirect.com/science/article/abs/pii/S0747563210002347>, zugegriffen am 19.05.2025.

Fahrmeir et. al. (2023). Statistik: Der Weg zur Datenanalyse. Springer-Verlag GmbH, DE.

- Gligorea, I. et. al. (2023). Adaptive Learning Using Artificial Intelligence in e-Learning: A Literature Review. <https://www.mdpi.com/2227-7102/13/12/1216>, zugegriffen am 20.05.2025.
- Gumprecht, S. & Meyer, S. & Peppel L. (2024). Lehren, Lernen und Schreiben mit generativen KI-Anwendungen: Empfehlungen und Impulse. [https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren\\_Lernen\\_und\\_Schreiben\\_mit\\_generativen\\_KIAAnwendungen\\_BeSt\\_BiB\\_Schreibwerkstatt\\_V4.2.pdf](https://www.frankfurt-university.de/fileadmin/standard/Hochschule/Bibliothek/Lehren_Lernen_und_Schreiben_mit_generativen_KIAAnwendungen_BeSt_BiB_Schreibwerkstatt_V4.2.pdf), zugegriffen am 14.06.2025.
- Hu, S. (2024). The Effect of Artificial Intelligence-Assisted Personalized Learning on Student Learning Outcomes: A Meta-Analysis Based on 31 Empirical Research Papers. *Science Insights Education Frontiers*, 24, 3873–3894. <https://doi.org/10.15354/sief.24.re395>
- Kem, D. (2022). Personalised and Adaptive Learning: Emerging Learning Platforms in the Era of Digital and Smart Learning. [https://www.researchgate.net/profile/Lukman-Prakoso/publication/359005289\\_feb\\_issue\\_soft\\_copy\\_2022/links/6221ef309f7b3246340dc43e/feb-issue-soft-copy-2022.pdf#page=26](https://www.researchgate.net/profile/Lukman-Prakoso/publication/359005289_feb_issue_soft_copy_2022/links/6221ef309f7b3246340dc43e/feb-issue-soft-copy-2022.pdf#page=26), zugegriffen am 19.05.2025.
- Kieslich, K., Lünich, M., Marcinkowski, F., & Starke, C. (2019). Hochschule der Zukunft—Einstellungen von Studierenden gegenüber Künstlicher Intelligenz an der Hochschule.
- Lackes, R. & Siepermann, M. (o. J.). Künstliche Intelligenz (KI). <https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-ki-40285/version-263673>, zugegriffen am 22.05.2025.
- Lange, T. & Mosler, K. (2017). Statistik kompakt: Basiswissen für Ökonomen und Ingenieure. Springer-Verlag Berlin Heidelberg.
- Lupa-Wojcik, I. (2024). Students' Willingness to Pay for Access to ChatGPT. *EUROPEAN RESEARCH STUDIES JOURNAL*, XXVII(Issue 3), 730–745. <https://doi.org/10.35808/ersj/3462>
- matplotlib (o. J.). Matplotlib: Visualization with Python. <https://matplotlib.org/>, zugegriffen am 14.06.2025.
- Merino-Campos, C. (2025). The Impact of Artificial Intelligence on Personalized Learning in Higher Education: A Systematic Review. *Trends in Higher Education*, 4, 17. <https://doi.org/10.3390/higheredu4020017>
- pandas (o. J.). pandas.DataFrame. <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>, zugegriffen am 14.06.2025.
- Scheiter et. al. (2025). Künstliche Intelligenz in der Schule: Eine Handreichung zum Stand in Wissenschaft und Praxis. [https://www.empirische-bildungsforschung-bmbf.de/img/KI\\_Review\\_20250318\\_Veroeffentlichung.pdf](https://www.empirische-bildungsforschung-bmbf.de/img/KI_Review_20250318_Veroeffentlichung.pdf), zugegriffen am 14.06.2025.
- Schwaab, M. & Durian, A. (Hrsg.) (2025). Zeitarbeit und Personaldienstleistungen – Chancen-Erfahrungen-Herausforderungen. Springer Fachmedien Wiesbaden GmbH.
- scipy (o. J.). Scipy: Fundamental algorithms for scientific computing in Python. <https://scipy.org/>, zugegriffen am 14.06.2025.
- seaborn (o. J.). An introduction to seaborn. <https://seaborn.pydata.org/tutorial/introduction.html>, zugegriffen am 14.06.2025.
- Streicher, A. (o. J.). Adaptive Lernsysteme. <https://www.iosb.fraunhofer.de/de/kompetenzen/systemtechnik/interoperabilitaet-assistenzsysteme/forschungsthemen/adaptive-lernsysteme.html>, zugegriffen am 22.05.2025.
- Zendler, A. (2018). Lerneffektivität ausgewählter Methoden. In: Zendler, A. (eds) Unterrichtsmethoden für den Informatikunterricht. Springer Vieweg, Wiesbaden. [https://link.springer.com/chapter/10.1007/978-3-658-20675-8\\_6](https://link.springer.com/chapter/10.1007/978-3-658-20675-8_6), zugegriffen am 22.06.2025

## **9 Kernfazit dieser Veröffentlichung**

Die vorliegende Arbeit vereint mehrere wissenschaftliche Beiträge, die sich mit dem Einsatz generativer KI in der Hochschulbildung und deren Auswirkungen auf verschiedene Aspekte des Lernens und Prüfens befassen. Die Analyse zeigt, dass Studierende generative KI-Tools wie ChatGPT überwiegend als unterstützend in der Prüfungsvorbereitung wahrnehmen, jedoch Bedenken hinsichtlich der Validität und akademischen Integrität äußern. Der Einfluss von KI auf Prüfungsangst bleibt unklar, da keine signifikanten Zusammenhänge festgestellt werden konnten.

Die Akzeptanz von KI-Tutoren wird als hoch eingeschätzt, wobei diese Technologien eher als Ergänzung zu traditionellen Lernmethoden betrachtet werden. Die Zahlungsbereitschaft für KI-gestützte Systeme ist gering, was die Notwendigkeit für Hochschulen unterstreicht, gezielte Schulungsangebote zu entwickeln.

Insgesamt zeigt die Arbeit, dass die Integration von KI in die Hochschulbildung das Potenzial hat, Lernprozesse zu revolutionieren, jedoch ein verantwortungsvoller Umgang und die Entwicklung neuer Kompetenzen erforderlich sind, um die Vorteile von KI-Technologien optimal zu nutzen und gleichzeitig die wissenschaftliche Integrität zu wahren.

**10****Bisher erschienene Working Papers des Fachbereich 3**

- Nr. 37 Michalski, Tino; Sohlbach, Simon (2025): Integration der wissenschaftlichen New Work- und Corporate Entrepreneurship-Ansätze im Hochschul- und Unternehmenskontext. DOI: <https://doi.org/10.48718/k9g9-c341>
- Nr. 36 Michalski, Tino; Sohlbach, Simon; Baroutas, Georgios (2025): Reshoring und De-Globalisierung: Relevanz der strategischen Entscheidung für das Business Development deutscher Unternehmen. DOI: <https://doi.org/10.48718/wfe6-ga57>
- Nr. 35 Michalski, Tino: (2025): Innovatives Business Development und Corporate Start-ups in nachhaltigen High Tech und High Service Clustern. DOI: <https://doi.org/10.48718/pbvr-tp84>
- Nr. 34 Lämmlein, Barbara; Lutz-Vock, Hannah; Engelmann, Sabrina: (2025): Hochschuldidaktik-Zertifikate im deutschen Bildungswesen – ein Vergleich. DOI: <https://doi.org/10.48718/gjwf-hd94>
- Nr. 33 Demiröz, Verena; Lämmlein, Barbara (2024): Data Literacy im Hochschulkontext. DOI: <https://doi.org/10.48718/ys2m-ww35>
- Nr. 32 Klesel, Michael; Messer, Uwe (2024): Substantive Use of Artificial Intelligence: The Role of Individual Differences. DOI: <https://doi.org/10.48718/8d9d-b049>
- Nr. 31 Bülbül, Dilek (2024): International Cooperation in Research: Cooperative Doctorates in Finance. DOI: <https://doi.org/10.48718/gyx4-g297>
- Nr. 30 Krause, Tobias; Ivanov, Igor; Sidki, Marcus (2023): Blame or Gain? Is Institutional Trust impacted by the Perception of Political In-fluence in State-owned Enterprises? DOI: <https://doi.org/10.48718/p69v-1y25>
- Nr. 29 Anderie, Lutz (2023): NFTs (Non-Fungible Tokens) – Funktion und Potenzialanalyse. DOI: <https://doi.org/10.48718/k9ak-b724>
- Nr. 28 Rosenbusch, Christoph et al. (2023): Building a European University Consortium: the Case of the U!REKA-Network. DOI: <https://doi.org/10.48718/je0y-9q78>
- Nr. 27 Anderie, Lutz; Höning, Michaela (2023): Untersuchungen zum Potenzial von Metaverse. DOI: <https://doi.org/10.48718/6xxa-c637>
- Nr. 26 Lämmlein, Barbara; Gerdiken, Ulrike (2002): Von Bigband bis Urban Gardening. Motivationale Gründe Studierender für ein kulturelles Engagement an Hochschulen. DOI: <https://doi.org/10.48718/twym-pw08>
- Nr. 25 Jung, Constantin (2022): Sustainable Corporate Governance in the United Kingdom. Environmental Sustainability in Directors' Decision-Making. DOI: <https://doi.org/10.48718/1k89-pj62>
- Nr. 24 Schlegler, Maren; Koch, Susanne (2022): Pilot study on the learning success of students in service-learning compared to other teaching and learning formats. DOI <https://doi.org/10.48718/n9fy-cv17>
- Nr. 23 Hagen, Tobias; Hamann, Jonas; Saki, Siavash (2022): Discretization of Urban Areas using POI-based Tesselation. DOI <https://doi.org/10.48718/7jjr-1c66>

- Nr. 22 Balioamoune, Mina; Bausony, Mohamed A.K.; Lutz, Stefan; K.A. Mohamed, Ehab (2022): International Ownership and SMEs in Middle Eastern and African Economies. DOI: <https://doi.org/10.48718/c5sy-3h14>
- Nr. 21 Graf, Erika; Franz, Catharina; Rugbarth, Matthias; Schmidt, Leonard (2021): Wie beeinflussen die Namen von Fleischersatzprodukten die Akzeptanz der Verbraucher? DOI: <https://doi.org/10.48718.smsh-p640>
- Nr. 20 Hagen, Tobias; Saki, Siavash; Scheel-Kopeinig, Sabine (2021): start2park – Determining, Explaining and Predicting Cruising for Parking.
- Nr. 19 Graf, Erika (2021): Banken auf dem Holzweg? Eine empirische Untersuchung der Bewertung von Kreditkarten aus Holz.
- Nr. 18 Ziegler, Yvonne; Uli, Vincenzo Uli; Kramer, Astrid; Tatari, Mahmoud Tatari (2021): Development of an innovative halal logistics concept for the air cargo supply chain.
- Nr. 17 Hagen, Tobias; Scheel-Kopeinig, Sabine (2020): Would Customers be willing to use an alternative (chargeable) delivery concept for the last mile?
- Nr. 16 Giegler, Nicolas; Schneider, Swen (2020): Leadership und Digitalisierung.
- Nr. 15 Graml, Regine; Hagen, Tobias; Ziegler, Yvonne; Khachatryan, Kristine; Astrida Herman, Ricky (2020): Lesbische Frauen in der Arbeitswelt – The L-Word in Business.
- Nr. 14 Voigt, Martina; Ruppert, Andrea (2018): Follow-up –Studie: Gendertypische Verhandlungskompetenz und ihre Auswirkungen auf Gehalts-und Aufstiegsverhandlungen.
- Nr. 13 Celebi, Kaan; Höning, Michaela (2018): Dynamic Macroeconomic Effects on the German Stock Market before and after the Financial Crisis.
- Nr. 12 Lutz, Stefan (2018): R&D, IP, and firm profits in the North American automotive supplier industry.
- Nr. 11 Voigt, Martina; Ruppert, Andrea (2016): Durchsetzungsorientiert, hart oder sachorientiert, kooperativ verhandeln – welche Einschätzungen und Erwartungen haben Studierende?
- Nr. 10 Hagen, Tobias (2016): Econometric Evaluation of a Placement Coaching Program for Recipients of Disability Insurance Benefits in Switzerland.
- Nr. 9 Weissenrieder, Caprice Oona; Spura, Anastassja (2015): Akzeptanz von Führungskräften- Analyse wahrgenommener Verhaltensweisen von Frauen und Männern in Führungspositionen.
- Nr. 8 Graf, Erika (2015): Raising Sustainability Awareness and Understanding in Higher Education.
- Nr. 7 Weissenrieder, Caprice Oona; Graml, Regine; Hagen, Tobias; Ziegler, Yvonne (2015): Explorative Untersuchung der Unternehmenskultur auf die Karrierechancen von Frauen.
- Nr. 6 Ruppert, Andrea; Voigt, Martina (2014): Verhandlungsstrategien und Verhandlungstaktiken in Gehaltsverhandlungen.
- Nr. 5 Jungmittag, Andre (2014): Combination of Forecasts across Estimation Windows: An Application to Air Travel Demand.

- Nr. 4 Jungmittag, Andre (2014): Der Trade-off zwischen deutschen Direktinvestitionen und Exporten: Wie wichtig sind die verschiedenen Dimensionen der Distanz?
- Nr. 3 Hagen, Tobias; Waldeck, Stefanie (2014): Using Panel Econometric Methods to Estimate the Effect of Milk Consumption on the Mortality Rate of Prostate and Ovarian Cancer.
- Nr. 2 Hagen, Tobias (2014): Impact of National Financial Regulation on Macroeconomic and Fiscal Performance After the 2007 Financial Shock Econometric Analyses Based on Cross-Country Data.
- Nr. 1 Rieck, Christian; Bendig, Helena; Hünnemeyer, Julius; Nitzsche, Lisa (2012): Diversität im Aufsichtsrat. Studie über die Zusammensetzung deutscher Aufsichtsräte.

Alle Publikationen des Fachbereich 3 Wirtschaft und Recht finden Sie unter:

<https://www.frankfurt-university.de/de/hochschule/fachbereich-3-wirtschaft-und-recht/forschung-und-transfer/publikationen/>