



Medienimpulse  
ISSN 2307-3187  
Jg. 63, Nr. 1, 2025  
doi: 10.21243/mi-01-25-11  
Lizenz: CC-BY-NC-ND-3.0-AT

# KI oder nicht KI? Textgeneratoren als Unterstützung des (wissenschaftlichen) Schreibprozesses

Karin Tengler

Gerhard Brandhofer

*Generative Künstliche Intelligenz (KI) ist in der Hochschulbildung angekommen und dient zunehmend der Nutzung im wissenschaftlichen Schreibprozess. Dieser Beitrag präsentiert eine Studie, die den Unterschied KI-generierter Texte versus von Lehramtsstudierenden (n=39) generierter Texte und den Nutzen generativer KI beim (wissenschaftlichen) Schreibprozess in der Hochschullehre untersucht. Die Ergebnisse der kriteriengeleiteten Bewertung zeigen, dass KI-Texte vor allem in den Bereichen des logischen Aufbaus und Sprachbeherrschung signifikant besser abschneiden. Außerdem wird vom Großteil der Studierenden generative KI nutzbringend im Schreibprozesses*

*wahrgenommen. Die Erkenntnisse lassen das Potenzial von KI als Werkzeug in der Hochschullehre erkennen, werfen jedoch auch Fragen hinsichtlich akademischer Integrität und kritischen Denkens auf.*

*Generative artificial intelligence (AI) has arrived in higher education and is increasingly being used in the academic writing process. This paper presents a study that examines the differences between AI-generated texts and texts produced by teacher education students (n=39), as well as the benefits of generative AI in the (academic) writing process within higher education. The results of the criterion-based evaluation show that AI-generated texts perform significantly better, particularly in terms of logical structure and language proficiency. Additionally, the majority of students perceive generative AI as a useful tool in the writing process. The findings highlight the potential of AI as a tool in higher education but also raise questions regarding academic integrity and critical thinking.*

## 1. Einleitung

Generative Künstliche Intelligenz (KI) hat sich mittlerweile als zunehmend verlässlicher Partner für Studierende etabliert. Laut einer deutschen Studie haben 98 % der Studierenden schon von generativer KI gehört und 78 % nutzen sie sowohl für schulische als auch für private Zwecke (Schlude et al. 2024), österreichische Studierende nutzen vor allem generative KI-Anwendungen aber auch verarbeitende KI, als Verwendungszweck werden vor allem Sprachverarbeitung, Informationssuche und Recherche sowie Textanalyse angegeben (Tulis-Oswald et al. 2024). Aktuelle Akzeptanzstudien, wie z. B. die von Brandhofer und Tengler (2024), zei-

gen aber auch, dass nicht nur Studierende sondern auch Lehrende dem Einsatz generativer KI grundsätzlich sehr positiv gegenüberstehen. Derzeit zählt ChatGPT von OpenAI, neben Google Gemini und Microsoft Copilot, zu den bekanntesten KI-Anwendungen, die unter anderem menschenähnliche Texte generieren können (Rahman/Watanobe 2023). Die rasche Verbreitung und der einfache Zugang zu den KI-gestützten Sprachmodellen machten schnell deutlich, dass sie nicht nur weitreichende Auswirkungen auf zahlreiche Lebensbereiche haben, sondern auch signifikante Veränderungen im Bildungssystem bewirken würden, das sich nach der Corona-Pandemie erneut einem bedeutenden Wandel gegenüber sieht (Albrecht 2023). Künstliche Intelligenz (KI) ist in vielen Anwendungsbereichen im täglichen Leben der Menschen bewusst oder auch unbewusst präsent. KI findet man nicht nur bei Text- und Bildgeneratoren, sondern unter anderem auch bei Technologien wie der automatischen Sprach- und Gesichtserkennung, bei personalisierten Feeds sozialer Medien sowie bei individuell zugeschnittenen Musik- und Kaufempfehlungen (Schirmer et al. 2023). KI-Text- und Bildgeneratoren sind in ihrer Basisversion häufig kostenfrei zugänglich, zeichnen sich durch hohe Leistungsfähigkeit und Benutzerfreundlichkeit aus und haben sich dadurch fest in der Gesellschaft etabliert (Wampfler 2023). Gerade die einfache Verfügbarkeit birgt aber auch die Gefahr, dass Studierende vermehrt KI-generierte Inhalte übernehmen, ohne sich intensiv mit einem Thema auseinanderzusetzen. KI-gestützte Anwendungen werden dabei nicht nur für das Verfassen von Aufsätzen genutzt, sondern auch zur Beantwortung von Fragen, zur Lö-

sung von Problemen, zur Erklärung komplexer Themen, für virtuelle Nachhilfe und zum Sprachenlernen (Rahman/Watanobe 2023). Das Potenzial der generativen KI ermöglicht es, Texte zu generieren, die von menschlich geschriebenen kaum zu unterscheiden sind. Dieser Beitrag präsentiert eine Studie, die untersucht, inwieweit mit KI-Tools generierte Texte als solche erkannt werden und inwieweit sie sich in ihrer Qualität von Menschen verfassten Texten unterscheiden. Außerdem wird der Frage nachgegangen, in welchen Bereichen des Schreibprozesses der Einsatz von generativer KI sinnvoll ist.

## 2. Theoretische und empirische Grundlagen

### 2.1 Einsatz generativer KI im Schreibprozess

Die derzeit verwendeten Systeme basieren auf selbstlernenden Sprachmodellen, sogenannten Large Language Models (LLMs), die durch das Training mit umfangreichen Textdaten ihre Mustererkennungsfähigkeiten kontinuierlich erweitern können. Anfragen von Nutzer\*innen werden so effektiv beantwortet, dass die generierten Texte für das Erzeugen sämtlicher Textsorten eingesetzt werden können:

Ihre Geschwindigkeit und die formale Korrektheit der automatisch generierten Texte helfen vielen Menschen, belastende Schreibarbeiten zu bewältigen. (Wampfler 2023: 2)

Rahman und Watanobe (2023: 18) beschreiben ChatGPT als ein revolutionäres Sprachmodell, das menschenähnliche Unterhaltungen führen und menschenähnlichen Text für jede Anfrage ge-

nerieren kann, der nahezu nicht zu unterscheiden ist. Neben der vielen Vorteile, die generative KI gerade für Studierende bringt, erwachsen durch sie auch Herausforderungen im Bildungsbereich. Da Textgeneratoren in der Lage sind, wissenschaftliche Texte zu allen möglichen Themen zu erzeugen, die menschenähnlich wirken, müssen sich Lehrkräfte und Bildungseinrichtungen auch potenzieller Betrugsrisiken bei schriftlichen Arbeiten und Prüfungen bewusst sein. Ebenso betonen Schirmer et al. (2023), dass die Implementierung von KI-Werkzeugen in pädagogischen Kontexten eine unvermeidliche Entwicklung darstellt, welche vielfältige pädagogische und methodische Herausforderungen sowie Möglichkeiten bietet. Zur effektiven Nutzung generativer KI-Technologien im Bildungswesen ist die Entwicklung und Implementierung spezifischer Formate sowie in der Aus-, Fort- und Weiterbildung erforderlich (Jeon/Lee 2023). Stolpe und Hallström (2024) erweitern diese Diskussion, indem sie die Bedeutung von Konzeptwissen, prozeduralem Wissen und kontextuellem Wissen für ein umfassendes Verständnis von KI in der Bildung beschreiben.

Eine grundlegende Unterstützung durch Textgeneratoren beim Verfassen von Texten erfolgt auf syntaktischer Ebene, etwa durch das Erkennen und Korrigieren von Tippfehlern. Auf semantischer Ebene können diese Modelle grammatikalische Inkonsistenzen identifizieren und mögliche Verbesserungsstrategien aufzeigen. Zusätzlich bieten sie die Möglichkeit, stilistische Verbesserungen themenspezifisch zu realisieren (Jeon/Lee 2023; Kasneci et al. 2023). Seit über zwei Jahrzehnten findet die Korrektur von Recht-

schreib- und Grammatikfehlern in Textverarbeitungsprogrammen mithilfe einfacher Datenbanken, die Wörter und Sätze abgleichen, automatisiert statt, mitunter auch nicht fehlerfrei. Durch den Einsatz generativer KI war, neben dem problemlosen Zugang und der vereinfachten Handhabung von KI-Textgeneratoren, auch ein signifikanter Qualitätsanstieg in den Ergebnissen der Autokorrektur seit einiger Zeit zu beobachten (Wampfler 2023).

Das große Potenzial der LLMs, wie ChatGPT, „liegt in der Fähigkeit natürliche Sprache zu verstehen und (schriftlich) zu produzieren“ (Schirmer et al. 2023: 5). Dies zeigt sich darin, dass sie auf Eingaben in Form von Text, Bild- oder Audiodateien mit schriftlichem Output reagieren. Folglich erweisen sich Technologien wie ChatGPT als wertvolle Werkzeuge im Schreibprozess, beispielsweise beim Verfassen von Essays (Malik et al. 2023). Der Schreibprozess selbst ist komplex, dynamisch und nicht-linear. Die Schreibforschung identifiziert drei grundlegende Subprozesse: Planen, Verschriften und Überarbeiten (Sieberer 2020). In all diesen Phasen können Textgeneratoren Unterstützung bieten, vom Finden entsprechender Ideen und Argumente über die Strukturierung des Textes bis hin zur Fehlerkorrektur, aber auch indem alternative Satzstrukturen zur Verbesserung der Klarheit und Kohärenz der Arbeit vorgeschlagen werden (Malik et al. 2023). Die schreibdidaktische Begleitung und Reflexion dieser Prozesse sind essenziell, um den Einsatz von KI-Tools sinnvoll zu evaluieren und zu entscheiden, wann ihre Verwendung angebracht ist (Wampfler 2023).

## 2.2 KI im Schreibprozess als Gegenstand der Forschung

Das Schreiben als Prozess zu begreifen, stellt für Schreibende eine erhebliche Erleichterung dar, indem der Schreibvorgang in aufeinander aufbauende Teilschritte zerlegt wird, die systematisch bearbeitet werden können, und somit die Komplexität des gesamten Prozesses deutlich reduziert wird. Wie im vorigen Kapitel erwähnt, lässt sich der Schreibprozess in drei Phasen unterteilen: die Vorbereitung, das Verfassen eines Roh texts und die anschließende Überarbeitung. Diese Phasen sind zyklisch und wiederholen sich oft mehrfach, bis der endgültige Text fertiggestellt ist. Bedeutend ist beim Prozessansatz, dass das Schreiben ein nicht-linearer Prozess ist, bei dem der Text kontinuierlich verbessert wird. Schreibende können jederzeit zwischen den einzelnen Phasen wechseln und Anpassungen vornehmen (Vedral/Ederer-Fick 2015: 224). Mit dem rasanten Anstieg der Verwendung generativer KI zur Unterstützung des Schreibprozesses wurde diese vermehrt auch zum Gegenstand der Forschung. Von Garrel et al. (2023: 10) identifizieren verschiedene effektive Anwendungsmöglichkeiten im Rahmen des Schreibprozesses von KI-basierten Werkzeugen, darunter die Durchführung von Literaturrecherchen, Übersetzungen in mehrere Sprachen, die Klärung von Verständnisfragen sowie die Entwicklung von Konzepten.

Darüber hinaus kann generative KI auf unerforschte Aspekte und aktuelle Forschungsthemen hinweisen, was Studierenden an Universitäten und Hochschulen helfen kann, Themen besser zu verstehen und zu analysieren und durch das reflektierende Anwen-

den auch zur Entwicklung und Förderung von kritischem Denken und Problemlösungsfähigkeiten beitragen (Kasneci et al. 2023). Studierende stehen dem Einsatz generativer KI in der Hochschulbildung äußerst positiv gegenüber und schätzen deren Unterstützung beim wissenschaftlichen Schreiben. Dies bestätigt unter anderem die Studie von Chan und Hu (2023). Dieser Aspekt wurde auch von Malik et al. (2023) in ihrer Studie, in der es um das Verfassen wissenschaftlicher Essays ging, hervorgehoben, die über den potenziellen Nutzen von generativer KI zur Förderung der Schreib Kreativität und zur Erleichterung der Erstellung origineller Kunstwerke und Literatur berichten.

Ein zentraler Forschungsstrang widmet sich der Qualität und Kohärenz KI-generierter wissenschaftlicher Texte im Vergleich zu menschlichen Arbeiten. Hörmann et al. (2024) untersuchten die Fähigkeiten von KI-generierten im Vergleich zu menschlich verfassten wissenschaftlichen Texten. Die Studie bekräftigt, dass KI-Anwendungen wie ChatGPT und Jenni AI strukturell solide und kohärente Texte erstellen können, betont jedoch, dass diese derzeit eher als Ergänzung, denn als Ersatz für menschlichen Intellekt und Kreativität dienen. Obwohl generative KI beim Verfassen von Gliederungen und Entwürfen effizient ist, fehlt es ihr oft an Tiefe und kritischer Analyse. Die Studie von Amirjalili et al. (2024), die darauf abzielte, KI-generierte wissenschaftliche Essays zu erkennen und dazu notwendige Kompetenzen zu identifizieren, zeigt, dass Lehrende oft auf fehlerorientierte Bewertungsmethoden zurückgreifen, wobei bei der Bewertung eher technische und gram-



matikalische Perfektion als Hinweise auf KI-Arbeiten gesehen werden und das Überprüfen der Richtigkeit von Fakten und der richtigen Quellenangaben vernachlässigt wurde. Die Studien belegen, wie wichtig es ist, eigenständiges Denken zu forcieren, institutionelle Richtlinien und Praktiken zu überprüfen oder anzupassen, um die akademische Integrität zu wahren und eine qualitativ hochwertige Hochschulbildung zu gewährleisten.

Die direkten Effekte der Nutzung generativer KI auf das wissenschaftliche Schreiben von Studierenden wurden in mehreren Studien beleuchtet. Lindqvist und Arvidsson (2024) untersuchten die Funktion von Reflexionen als pädagogisches Instrument und stellten die von KI (ChatGPT) generierten den von Studierenden verfassten Reflexionstexte gegenüber. Die Ergebnisse zeigten, dass die studentischen Texte persönliche Einsichten und Methoden zur Erleichterung des Lernens enthielten, während die KI-Texte überwiegend umfassende Beantwortung der Reflexionsfragen aufwiesen. Die KI-generierten Texte stellten keine Bedrohung für die akademische Integrität dar, sondern boten vielmehr eine wertvolle Ergänzung, um die Studierenden bei der Intensivierung ihrer Reflexion zu unterstützen.

In der Studie von Basic et al. (2023) wurde an einer Gruppe von Masterstudierenden untersucht, inwiefern die Verwendung von ChatGPT-3 Einfluss auf das wissenschaftliche Schreiben der Studierenden hat. Fokus der Forschung waren die möglichen Effekte auf die Notenverbesserung, die Reduktion der Schreibzeit und die Veränderung der Textauthentizität. Obwohl ChatGPT in der Lage

ist, kontextuell relevante Inhalte zu generieren, zeigen die Ergebnisse der Studie, dass der Einsatz von ChatGPT-3 nicht zu einer signifikanten Steigerung der Aufsatzqualität führte. Die Gruppe, die ChatGPT-3 zur Unterstützung heranzog, erzielte keine besseren Ergebnisse hinsichtlich der Textinhalte und der Schnelligkeit, aber der Einsatz hatte positive Effekte auf die Ideenfindung und um die anfängliche Schreibphase zu beschleunigen. Als mögliche Erklärung wurde angeführt, dass die Studierenden mit dem Tool noch nicht so vertraut waren. Herbold et al. (2023) verglichen in ihrer Studie Essays, die von Schüler\*innen der Oberstufe verfasst wurden, mit jenen, die von ChatGPT-3 und ChatGPT-4 generiert wurden. Jede\*r der 139 teilnehmenden Lehrpersonen bewertete jeweils sechs Texte nach verschiedenen Kriterien, wie z. B. Thema und Vollständigkeit.

Die Studie zeigt, dass ChatGPT qualitativ hochwertige argumentative Aufsätze generieren kann und zudem die Qualität der von Menschen verfassten Essays in allen Bewertungsrubriken deutlich übertrifft. Zu erkennen war auch, dass ChatGPT-4 zusätzlich noch bessere Ergebnisse lieferte als die Vorgängervariante. Der Schreibstil generativer KI unterscheidet sich insofern, dass ChatGPT-Modelle mehr Nominalisierungen verwenden und die Sätze eine komplexere Satzkonstruktion aufweisen.

Die Fähigkeit generativer KI-Modelle, semantisch und syntaktisch korrekte Texte zu produzieren, führt zu einer tiefgreifenden Veränderung auch im Hochschulbereich, wie wissenschaftliche Texte erstellt und bewertet werden. Dies wirft wichtige Fragen für die

Forschung auf. Einen Beitrag zur Forschung will diese Studie mit der Beantwortung der folgenden Forschungsfragen leisten:

- Inwieweit gelingt es zu erkennen, ob ein Text von einem Menschen oder generativer KI verfasst wurde?
- Inwieweit unterscheiden sich KI-generierte Texte in sprachlicher und inhaltlicher Hinsicht von Texten, die von Menschen verfasst wurden?
- In welchen Bereichen des Schreibprozesses wird der Einsatz generativer KI von Studierenden als nutzbringend wahrgenommen?

### 3. Methodologie

#### 3.1 Teilnehmende und Setting

Im Rahmen des Bachelorstudiums Primarstufe im Studienjahr 2023/24 wurde in der Orientierung *Digitalität* die Lehrveranstaltung mit dem Titel „Kinder lernen digital (Lehren und Lernen mit digitalen Medien, die Gestaltung digitaler Lernräume)“ angeboten, die auf die Implementierung und Auseinandersetzung mit Themen der Medienpädagogik den Fokus legt und aus einer Vorlesung und einer Werkstatt besteht. Inhalt und Struktur der Lehrveranstaltung leiten sich aus dem Kompetenzmodell *digi.kompP* ab (Brandhofer et al. 2020). Studierende des zweiten Semesters nehmen daran verpflichtend teil.

In der Vorlesung erhielten die Studierenden eine Einführung in das Thema Künstliche Intelligenz, wobei der Fokus auf den Grundlagen der generativen KI und ihrem didaktischen Einsatz lag. Als Teil der dazugehörigen Werkstatt wurden die Studierenden, die in der Werkstatt in zwei Gruppen geteilt waren, beauftragt, jeweils ein Essay zum Thema „Digitale Medien in der Primarstufe“ im Umfang von 200 Wörtern selbst zu verfassen und zwei Essays von verschiedenen KI-Tools, wie ChatGPT, Google Gemini, Microsoft Copilot etc. im gleichen Umfang generieren zu lassen.

Der Prompt, der eingegeben werden sollte, lautete: Schreibe einen Aufsatz mit etwa 200 Wörtern über „Digitale Medien in der Primarstufe - Potenziale und Risiken“. Die Texte wurden von den Studierenden in ein Formular eingefügt, auf der Lernplattform hochgeladen und vom Forscher\*innenteam anonymisiert. Danach bekamen die Studierenden der jeweils anderen Gruppe zwei zufällig ausgewählte Texte, die zunächst als *KI-generiert oder nicht* eingeschätzt und nach vorgegebenen Kriterien bewertet wurden. Diese Bewertung wurde in einem Onlinefragebogen eingetragen.

Insgesamt nahmen 39 Lehramtsstudierende an der Erstellung und der Bewertung der Essays teil. Somit konnten 78 Texte, die sich aus 41 KI-generierten und 37 von Menschen verfassten Texten zusammensetzten, bewertet werden. Die Abgabe der Texte diente gleichzeitig als Einverständnis, sich an der Studie zu beteiligen. Eine Zuordnung der Teilnehmenden nach Geschlecht schien den Autor\*innen für diese Studie nicht relevant.

### 3.2 Fragebogen

Der Fragebogen, der mit Microsoft Forms erstellt wurde, bestand aus drei Abschnitten. Der erste Abschnitt beinhaltete sieben Kategorien, die in Anlehnung an die Studie von Herbold et al. (2023) verwendet wurden. Dazu gehören: Thema und Vollständigkeit, Logik und Komposition, Ausdruckskraft und Verständlichkeit, Beherrschung der Sprache, Komplexität der Satzstrukturen, Wortschatz und Textverknüpfung sowie Sprachkonstruktion. Für jede Kategorie wurde eine siebenstufige Likert-Skala mit Werten von null bis sechs festgelegt, wobei null der schlechteste Wert ist (z. B. keine logische Struktur) und sechs der beste Wert (z. B. klare Struktur und logische und fundierte Argumentation).

Die Studierenden erhielten eine Liste mit den Kategorien und einer Beschreibung der zugehörigen Bewertungsstufen von null bis sechs als Anleitung. Der zweite Abschnitt des Fragebogens beschäftigte sich mit der Einschätzung, ob der vorliegende Text von einem Menschen oder einer generativen KI erstellt wurde. Abschließend wurden die Studierenden zu ihrer Einstellung bezüglich der Eignung von KI-gestützten Textgeneratoren im Schreibprozess befragt. Die vorgeschlagenen Items entstammen der im Theorieteil genannten Studien (Schlude et al. 2024; von Garrel et al. 2023). Jedes Item konnte mit einer vierstufigen Likert-Skala von 4 (trifft zu) bis 1 (trifft nicht zu) bewertet werden.

### 3.3 Auswertung

Da alle Fragebögen vollständig ausgefüllt waren, konnten alle Datensätze für die Auswertung verwendet werden. Die statistische Analyse erfolgte mittels der Software IBM SPSS 29. Zunächst wurde der Mittelwert erhoben. Um einen möglichen Unterschied bei der Bewertung der Texte in den einzelnen Kategorien, sprachlicher und inhaltlicher Aspekte betreffend, festzustellen, wurde der t-Test, ein Test für den Vergleich gepaarter Stichproben (Döring 2023: 150), angewendet. Die Daten bezüglich des wahrgenommenen Nutzens generativer KI im Schreibprozess wurden mittels deskriptiver Statistik ausgewertet. Die Ergebnisse werden im folgenden Abschnitt vorgestellt und diskutiert.

## 4. Ergebnisse und Interpretation

Die Ergebnisse der Studie werden zunächst präsentiert, indem die Einschätzungen der Studierenden vorgestellt werden, ob ein Text von einer KI oder einem Menschen verfasst wurde. Anschließend wird die Beurteilung sprachlicher und inhaltlicher Aspekte der beiden Textgruppen dargelegt. Abschließend erfolgt die Analyse der Wahrnehmungen der Studierenden hinsichtlich der Effektivität generativer KI in verschiedenen Bereichen des Schreibprozesses.

### 4.1 KI oder nicht KI

Das Balkendiagramm (Abb. 1) zeigt die Ergebnisse der Einschätzung von Studierenden, ob die ihnen zur Bewertung vorgelegten

Essays mit oder ohne die Hilfe einer Künstlichen Intelligenz verfasst wurden.

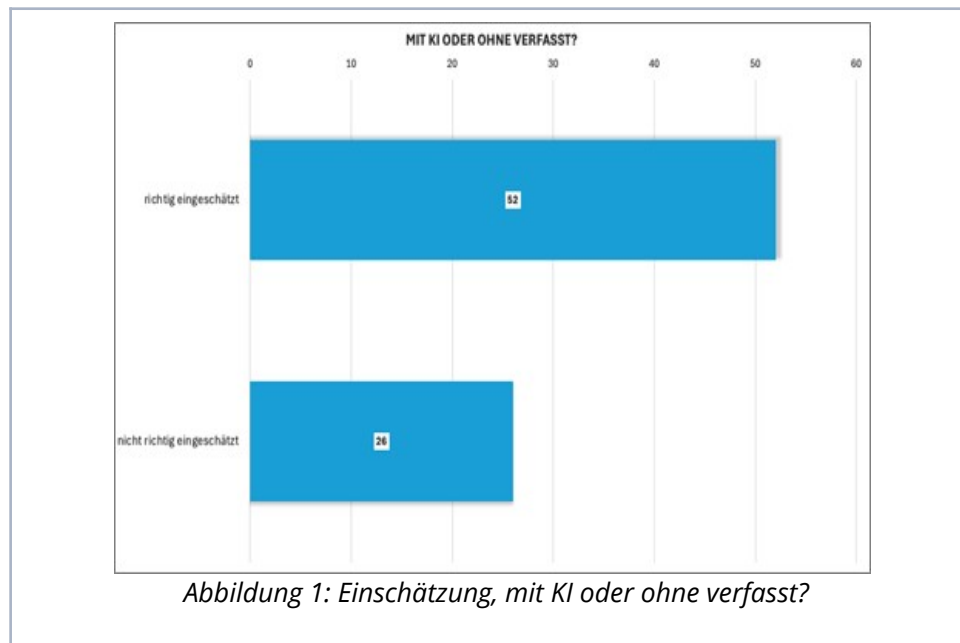


Abbildung 1: Einschätzung, mit KI oder ohne verfasst?

Von den 78 Bewertungen haben die Studierenden in 52 Fällen richtig eingeschätzt, ob ein Textgenerator bei der Erstellung des Essays verwendet wurde oder nicht. Das bedeutet, dass der größere Teil, immerhin genau zwei Drittel der Einschätzungen (66,67 %), die Arbeit generativer KI erkannte, aber immerhin doch ein Drittel auch danebenlag, was vermuten lässt, dass KI-verfasste Texte in vielen Fällen von menschlich verfassten Texten doch nicht einfach zu unterscheiden sind oder dass sie als Studierende aufgrund noch nicht ausreichender Erfahrung in der Textevaluati-on noch nicht vollständig in der Lage sind, die Unterschiede zu erkennen.

## 4.2 Bewertung der Texte

Die zweite Forschungsfrage zielte darauf ab, zu untersuchen, inwieweit sich KI-generierte Texte in sprachlicher und inhaltlicher Hinsicht von Texten, die von Menschen verfasst wurden, unterscheiden. Tabelle 1 zeigt die statistischen Ergebnisse nach Anwendung eines gepaarten t-Tests, der in der empirischen Forschung durchgeführt wird, um signifikante Unterschiede zwischen abhängigen oder unabhängigen Stichproben zu identifizieren (Döring 2023: 150). In diesem Fall ging es darum, die Bewertung der Texte durch die Studierenden in den Kategorien Thema und Vollständigkeit, Logik und Komposition, Ausdruckskraft und Verständlichkeit, Beherrschung der Sprache, Komplexität der Satzstrukturen, Wortschatz und Textverknüpfung sowie Sprachkonstruktion zu analysieren.

Die Effektstärken (Cohen's  $d$  und Hedge's  $g$ ), die angeben, wie groß der Unterschied zwischen den Gruppen ist (Döring 2023: 802), liegen bei den meisten Paaren im Bereich von kleinem bis mittlerem Effekt. Die Hedge's  $g$ -Korrektur wird angewendet, um eine genauere Schätzung der Effektgröße bei kleinen Stichproben zu erhalten.

*Kategorie Thema und Vollständigkeit:* Die sehr geringe Effektgröße (Cohen's  $d = 0,071$ ) und das 95 % Konfidenzintervall deuten darauf hin, dass es keinen statistisch signifikanten Unterschied in der Themenbehandlung und Vollständigkeit zwischen KI-generierten und menschlich verfassten Texten gibt.



Paired Samples Test												
Kategorie	Mean	Std. Dev.	Std. Error	95% Conf. I./Diff. Lower	95% Conf. I./Diff. Upper	t	df	One-Sided p	Two-Sided p	Cohen's d	Hedges' corr.	
Thema und Vollständigkeit	-,081	1,140	,187	-,299	,461	,433	36	,334	,668	,071	,070	
Logik und Komposition	,405	1,279	,210	-,021	,832	1,928	36	,031	,062	,317	,310	
Ausdrucksfähigkeit und Verständlichkeit	,162	1,555	,256	-,356	,681	,634	36	,265	,530	,104	,102	
Beherrschung der Sprache	,514	1,446	,238	,031	,996	2,161	36	,019	0,37	,355	,348	
Komplexität	,486	2,050	,337	-,197	1,170	1,444	36	,079	,157	,237	,232	
Wortschatz und Textverknüpfung	,432	1,772	,291	-,158	1,023	1,484	36	,073	,146	,244	,239	
Sprachkonstrukte	,189	1,613	,265	-,349	,727	,713	36	,240	,480	,117	,115	

*Tab. 1: Datenauswertung*

*Kategorie Logik und Komposition:* Die mittlere Effektgröße (Cohen's  $d = 0,317$ ) und das Konfidenzintervall lassen darauf schließen, dass KI-generierte Texte in Bezug auf einen logischen Aufbau besser abschneiden als menschlich verfasste Texte.

*Ausdrucksfähigkeit und Verständlichkeit:* Durch die geringe Effektgröße (Cohen's  $d = 0,104$ ) ist hier kein signifikanter Unterschied in der Ausdrucksfähigkeit und Verständlichkeit beider Textgruppen erkennbar.

*Beherrschung der Sprache:* Die moderate Effektgröße (Cohen's  $d = 0,355$ ) und das Konfidenzintervall könnten darauf hinweisen, dass KI-generierte Texte eine bessere Beherrschung der Sprache zeigen.

*Komplexität:* Trotz der moderaten Effektgröße (Cohen's  $d = 0,237$ ) ist das Konfidenzintervall nicht aussagekräftig genug, um einen signifikanten Unterschied in der Komplexität der Texte zu bestätigen.

*Wortschatz und Textverknüpfung:* Auch hier lässt die moderate Effektgröße (Cohen's  $d = 0,244$ ) mit einem Konfidenzintervall, das Null einschließt, keinen klaren Schluss auf signifikante Unterschiede im Wortschatz und in der Textverknüpfung zwischen den beiden Textgruppen zu.

*Sprachkonstrukt:* Die geringe Effektgröße (Cohen's  $d = 0,117$ ) weisen darauf hin, dass keine signifikanten Unterschiede der verwendeten Sprachkonstrukte zwischen KI-generierten und von Menschen verfassten Texten festgestellt wurden.

#### 4.3 Nutzung von Textgeneratoren

Im letzten Abschnitt des Fragebogens wurden die Studierenden befragt, bei welchen Phasen des Schreibprozesses der Einsatz generativer KI von Studierenden als nutzbringend wahrgenommen wird. Die Antworten wurden mittels einer vierstufigen Likert-Skala erfasst, deren Ergebnisse in Abbildung 2 dargestellt sind.

Kategorie	N	M	SD	Min.	Max.
Verfassen von Texten	39	2,92	0,900	1	4
Strukturieren, Konzepterstellung	39	3,38	0,633	2	4
Argumente finden	39	3,54	0,643	2	4
Präzises Formulieren	39	2,67	0,838	1	4
Lösen von Schreibblockaden	39	3,38	0,847	1	4
Fehlervermeidung	39	2,54	0,884	1	4

*Tabelle 2: Wahrgenommener Nutzen im Schreibprozess*

Die deskriptive Analyse der Datensätze über die Aspekte der Nutzung sind in Tabelle 2 zu finden.

Tabelle 2 liefert Einblicke in die Wahrnehmung von Studierenden darüber, wie nutzbringend generative KI als Unterstützung in den verschiedenen Phasen des Schreibprozesses ist. Nachfolgend findet man die Auswertung und Interpretation der Bewertung.

*Verfassen von Texten:* Mit einem Durchschnittswert von  $M = 2,92$  ( $SD = 0,900$ ) bewerten Studierende die Nützlichkeit der KI für das grundlegende Verfassen von Texten als moderat. Die hohe Standardabweichung lässt auf heterogene Meinungen hinweisen, was nahelegt, dass der empfundene Nutzen generativer KI bei der Erstellung von Essays von Person zu Person sehr kontrovers gesehen wird.

*Strukturieren, Konzepterstellung:* Textgeneratoren werden als Werkzeug zum Strukturieren und zum Erstellen von Konzepten ( $M = 3,38$ ,  $SD = 0,633$ ) als recht nutzbringend wahrgenommen. Dies spiegelt auch die Ergebnisse von KI-Studien wider (Hörmann et al. 2024; Schlude et al. 2024; von Garrel et al. 2023).

*Argumente finden:* Mit der höchsten durchschnittlichen Bewertung ( $M = 3,54$ ,  $SD = 0,643$ ) scheinen Studierende generative KI besonders effektiv zu finden, wenn es darum geht, Argumente für ihre Texte zu generieren. Die geringe Standardabweichung weist darauf hin, dass die Studierenden hier weitgehend einer Meinung sind. Dies deckt sich auch mit den Ergebnissen der Studien von Basic et al. (2023) und Malik et al. (2023).

*Präzises Formulieren:* Studierende bewerten die Fähigkeit der Textgeneratoren, beim präzisen Formulieren zu helfen, gesamt gesehen als weniger nützlich ( $M = 2,67$ ,  $SD = 838$ ). Die höhere Standardabweichung und ein Blick auf das Diagramm zeigen aber, dass die Antworten nicht sehr homogen ausfallen, denn beinahe die Hälfte hat KI als Werkzeug zum treffenden Formulieren dennoch als positiv wahrgenommen.

*Lösen von Schreibblockaden:* Auch hier wird der KI ein nutzbringender Aspekt zugesprochen ( $M = 3,38$ ,  $SD = 0,847$ ). Obwohl die Spannweite der Bewertungen und die Standardabweichung auf unterschiedliche Erfahrungen hinweisen, findet dennoch die Mehrheit der Studierenden, dass Textgeneratoren als sehr hilfreich beim Überwinden von Schreibblockaden gesehen werden

können. Dies deckt sich mit den Wahrnehmungen in der Studie von Basic et al. (2023).

*Fehlervermeidung:* Dieser Bereich schnitt bei dieser Befragung zwar am schlechtesten ab ( $M = 2,54$ ,  $SD = 0,884$ ), dennoch fällt die Bewertung bei mehr als der Hälfte der Befragten noch immer positiv aus. Möglicherweise liegt dies an der Formulierung des Items. Vielleicht wäre die Bezeichnung „zur Fehlerkorrektur“ treffender formuliert gewesen.

Insgesamt zeigen die Ergebnisse, dass in der Kategorie Logik und Komposition sowie auch im Bereich der Sprachbeherrschung signifikante Unterschiede gegeben sind, wobei KI-generierte Texte allgemein besser abschneiden. In anderen Bereichen wie Ausdrucksfähigkeit, Komplexität und Sprachkonstruktion sind die Unterschiede nicht signifikant. Dass nur in zwei Kategorien signifikante Unterschiede deutlich wurden, kann möglicherweise auch an der geringen Größe der Stichprobe liegen.

#### 4.4 Diskussion

Die Ergebnisse lassen den Schluss zu, dass die Qualität KI-generierter Texte im Vergleich zu jenen von Menschen verfasster Texte hinsichtlich inhaltlicher und sprachlicher Aspekte höher ausfällt, wobei die Texte vor allem in den Bereichen Logik und Vollständigkeit hervorstechen. Dies lässt sich mit der Studie von Hörmann et al. (2024) vergleichen, die die Fähigkeit von KI-Tools zur Erstellung strukturell kohärenter Texte mit einer logischen Struktur betont. Diese Erkenntnisse finden sich auch in der Studie von Herbold et

al. (2023), bei der die Texte, die von ChatGPT-3 und ChatGPT-4 generiert wurden, die Qualität der menschlich verfassten Texte übertrafen. Darüber hinaus zeigen unsere Ergebnisse, dass KI-Applikationen in korrekter Sprache ausgeben, wie die moderate Effektstärke in dieser Kategorie belegt. Dies entspricht den Beobachtungen von Kasneci et al. (2023), die darauf hinweisen, dass KI-Textgeneratoren grammatikalisch und orthografisch richtige Texte produzieren, wodurch sie Inkonsistenzen erkennen und potenzielle Verbesserungsstrategien vorschlagen können. Hinsichtlich der Ausdrucksstärke und Vollständigkeit von Texten zeigen die Ergebnisse keine statistisch signifikanten Unterschiede zwischen KI-generierten und von Menschen verfassten Texten. Das könnte bedeuten, dass KI-Tools in ihrer aktuellen Form möglicherweise nicht in der Lage sind, die in menschlichen Texten beobachtete, inhaltliche Tiefe vollständig zu erfassen oder nachzubilden. Dies könnte ein Bereich sein, in dem menschliche Autor\*innen weiterhin einen Vorteil behalten, wie die Ergebnisse von Muñoz-Ortiz et al. (2024) belegen, die auf eine größere Variabilität in deren Satzlängen und Wortschatz verweisen. Dieser Aspekt war in der vorliegenden Studie vermutlich aufgrund der relativ kurzen Textlängen nicht erkennbar. Die Untersuchung von Malik et al. (2023) zum Potenzial von KI im Bereich des wissenschaftlichen Schreibens und die Ergebnisse, die belegen, dass KI-Tools hilfreich sind, um die Schreib kreativität zu fördern und wissenschaftliche Essays zu verfassen, stützen die Erkenntnisse der Studie zur Nützlichkeit von generativer KI im Hinblick auf sprachliche Kompetenz und logische Struktur. Dies würde auch die Erkennt-

nisse von Lindqvist and Arvidsson (2024) bestätigen, die KI-Tools als wertvolle Ergänzung für den Schreibprozess sehen, um sachliche Formulierungen zu fördern.

Die Einschätzungen über den wahrgenommenen Nutzen lassen den Schluss zu, dass Textgeneratoren als überwiegend positiv zur Unterstützung in allen Phasen des Schreibprozesses gesehen werden. Dabei verwenden Studierende generative KI vorrangig als Werkzeug für die Ideenentwicklung und Argumentation, wobei das Potenzial für Konzipieren und Strukturieren von Texten hervorgehoben werden kann. Dagegen sind die Meinungen zur Effektivität der KI beim präzisen Formulieren und bei der Fehlervermeidung geteilt, was auf Limitationen in diesen Bereichen hinweisen könnte. Insgesamt legen die Ergebnisse der Auswertung nahe, dass generative KI im Schreibprozess überwiegend als hilfreiches und unterstützendes Werkzeug wahrgenommen wird. Diese Einschätzung deckt sich mit den Erkenntnissen aus Akzeptanzstudien, wie sie beispielsweise von Brandhofer und Tengler (2024) dargelegt wurden.

Die Ergebnisse der Studie zum Vergleich von KI-generierten und von Menschen verfassten Texten haben relevante Implikationen für die akademische Integrität und die Bildungspraxis. KI-Tools können den Schreibprozess durch strukturelle und grammatikalische Unterstützung erleichtern, bergen jedoch zugleich das Risiko unethischer Praktiken im wissenschaftlichen Arbeiten, wenn sie von Studierenden genutzt werden, um Aufgaben zu bearbeiten, die eigenständiges, kreatives oder analytisches Denken erfordern.

Dies könnte die Authentizität wissenschaftlicher Arbeiten herausfordern und die Integrität studentischer Leistungen beeinflussen, wodurch die Entwicklung neuer Richtlinien und die Anpassung von Prüfungsformaten sinnvoll erscheinen.

## 5. Fazit und Ausblick

Grundsätzlich wird mit dieser Studie das Potenzial generativer KI-Applikationen im Bildungsbereich bestätigt, insbesondere im (wissenschaftlichen) Schreibprozess. Die Ergebnisse zeigen, dass KI-generierte Texte in bestimmten Aspekten des wissenschaftlichen Schreibens, wie dem logischen Aufbau und der Sprachbeherrschung, signifikant besser abschneiden als von Menschen verfasste Texte. Die Einschätzungen der Studierenden belegen, dass KI-Tools insbesondere bei der Ideenfindung und Argumentation als wertvolle Unterstützung wahrgenommen werden und auch in weiteren Phasen des wissenschaftlichen Schreibens als hilfreich gelten. Bildungseinrichtungen stehen vor der Herausforderung, Lehrpläne so anzupassen, dass KI-Anwendungen das Lernen unterstützen, ohne kritisches Denken und individuelle Analysefähigkeit zu untergraben.

Um das Potenzial dieser Technologien verantwortungsvoll zu nutzen, sollten Bildungseinrichtungen die Integration von KI-Werkzeugen fördern, dabei jedoch den Erhalt der akademischen Integrität und die Förderung kritischen Denkens im Blick behalten. Besonders im Hochschulbereich umfasst Wissenschaftsethik nicht nur die korrekte Nutzung von Quellen, sondern auch die



Originalität der Argumentation und die eigenständige Entwicklung wissenschaftlicher Positionen. Der Einsatz von generativer KI birgt das Risiko, dass Studierende sich zu sehr auf automatisierte Inhalte stützen, ohne eigene Analyse- und Synthesefähigkeiten zu entwickeln. Zudem können KI-generierte Texte fehlerhafte oder erfundene Quellen enthalten, was eine sorgfältige Überprüfung und kritische Reflexion der generierten Inhalte erfordern. Daher sollten Studierende dazu ermutigt werden, KI nicht als Ersatz, sondern als unterstützendes Werkzeug für den wissenschaftlichen Schreibprozess zu nutzen, das ihnen hilft, eigene Ideen weiterzuentwickeln und ihre Argumentationsstrukturen zu schärfen. Angesichts der dynamischen Entwicklung und der vielfältigen Einsatzmöglichkeiten von KI-Tools ist es notwendig, dass Lehrkräfte und Studierende sowohl grundlegende Kompetenzen als auch ein kritisches Bewusstsein im Umgang mit diesen Technologien erwerben, um eine reflektierte und fundierte Nutzung generativer KI im Bildungsbereich zu ermöglichen.

Zukünftige Forschung könnte darauf abzielen, gezielte didaktische Strategien und Richtlinien zu entwickeln und zu untersuchen, die den integren Einsatz von KI im wissenschaftlichen Arbeiten sicherstellen.

## Literatur

Albrecht, Steffen (2023): ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen.

<https://doi.org/10.5445/IR/1000158070>.

Amirjalili, Faezeh/Neysani, Maryam/Nikbakht, Amir (2024): Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature, in: *Frontiers in Education* 9, 1347421.

Basic, Zelana/Banovac, Ana/Kruzic, Ivana/Jerkovic, Ivan (2023): Better by you, better than me, ChatGPT3 as writing assistance in students' essays, in: *arXiv – preprint online* unter: <https://arxiv.org/abs/2302.04536> (letzter Zugriff: 20.02.2025).

Brandhofer, Gerhard/Miglbauer, Marlene/Fikisz, Walter/Höfler, Elke/Kayali, Fares/Steiner, Michael/Prohaska, Julia/Riepl, Andreas (2020): Das digi.kompP Kompetenzmodell. Version Dezember 2019, Grafik und Deskriptoren. Onlinecampus Virtuelle PH im Auftrag des BMBWF, online unter:

[https://www.bmbwf.gv.at/Themen/schule/zrp/dibi/paed/digi\\_komp.html](https://www.bmbwf.gv.at/Themen/schule/zrp/dibi/paed/digi_komp.html) (letzter Zugriff: 20.02.2025).

Brandhofer, Gerhard/Tengler, Karin (2024): Zur Akzeptanz von KI-Applikationen bei Lehrenden und Lehramtsstudierenden, in: *R&E-SOURCE* 11(3), 7–25.

<https://doi.org/10.53349/resource.2024.i3.a1277>.

Chan, Cecilia Ka Yuk/Hu, Wenjie (2023): Students' voices on generative AI: Perceptions, benefits, and challenges in higher education, in: *International Journal of Educational Technology in Higher Education* 20(1), 43.

Döring, Nicola (2023): *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6. Aufl.), Wiesbaden: Springer.

Herbold, Steffen/Hautli-Janisz, Anette/Heuer, Ute/Kikteva, Zlatka/Trautsch, Alexander (2023): A large-scale comparison of human-written versus ChatGPT-generated essays, in: *Scientific Reports* 13(1), 18617.

Hörmann, Corinna/Kuka, Lisa/Fraser, Anneliese/Sabitzer, Barbara (2024): Innovation or Imitation? A Critical Analysis of AI-Authored vs. Human-Authored Scientific Papers, in: *CSEDU* (1), 534–541.

Jeon, Jungmin/Lee, Seungbin (2023): Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT, in: *Education and Information Technologies* 28(12), 15873–15892.

Kasneci, Enkelejda/Seßler, Kathrin/Küchemann, Stefan/Bannert, Maria/Dementieva, Daryna/Fischer, Frank/Gasser, Urs/Groh, Georg/Günemann, Stephan/Hüllermeier, Eyke (2023): ChatGPT for good? On opportunities and challenges of large language models for education, in: *Learning and Individual Differences* 103, 102274.

Malik, Agung Rinaldy/Pratiwi, Yuni/Andajani, Kusubakti/Numer-tayasa/I Wayan Numertayasa/Suharti, Sri/Darwis, Arisa (2023): Exploring Artificial Intelligence in Academic Essay: Higher Education Student's Perspective, in: International Journal of Educational Research Open 5, 100296.

Rahman, Md. Mostafizer/Watanobe, Yutaka (2023): ChatGPT for education and research: Opportunities, threats, and strategies, in: Applied Sciences 13(9), 5783.

Schirmer, Katja/Berger, Martin/Himpsl-Gutermann, Klaus/Lorenz, Setera-Anna/Steiner, Michael (2023): Künstliche Intelligenz im Unterricht: Lehr-/Lernszenarien für verschiedene Gegenstände, in: Medienimpulse 61(2), 40 Seiten, online unter: <https://journals.univie.ac.at/index.php/mp/article/view/7999> (letzter Zugriff: 20.02.2025).

Schlude, Antonia/Mendel, Ulrike/Stürz, Roland/Fischer, Micha (2024): Verbreitung und Akzeptanz generativer KI an Schulen und Hochschulen, online unter: <https://www.bidt.digital/publikation/verbreitung-und-akzeptanz-generativer-ki-an-schulen-und-hochschulen/> (letzter Zugriff: 20.02.2025).

Sieberer, Elisabeth (2020): Schreibkompetenz vermitteln: Ein Blick auf die Unterrichtspraxis, in: R&E-SOURCE.

Stolpe, Karin/Hallström, Jonas (2024): Artificial intelligence literacy for technology education, in: Computers and Education Open 6, 100159.

Tulis-Oswald, Maria/Kinskofer, Franziska/Fischer, Elena (2024): Ergebnisse der quantitativen Erhebung zur KI-Nutzung an Hochschulen, in: Brandhofer, Gerhard/Gröblinger, Ortrun/Jadin, Tanja/Raunig, Michael/Schindler, Julia (Hg.): Von KI lernen, mit KI lehren: Die Zukunft der Hochschulbildung, 76–125, BoD, online unter: <https://fnma.at/projekte/vereinsprojekte/von-ki-lernen-mit-ki-lehren-die-zukunft-der-hochschulbildung> (letzter Zugriff: 20.02.2025).

Vedral, Johanna/Ederer-Fick, Elfriede (2015): Schreibforschung und Schreibdidaktik: Modelle und Theorien wissenschaftlichen Schreibens, in: Egger, Rudolf/Wustmann, Cornelia/Karber, Anke (Hg.): Forschungsgeleitete Lehre in einem Massenstudium: Bedingungen und Möglichkeiten in den Erziehungs- und Bildungswissenschaften, Wiesbaden: Springer, 217–238.

von Garrel, Jörg/Mayer, Jana/Mühlfeld, Markus (2023): Künstliche Intelligenz im Studium: Eine quantitative Befragung von Studierenden zur Nutzung von ChatGPT & Co. [https://doi.org/10.48444/h\\_docs-pub-395](https://doi.org/10.48444/h_docs-pub-395).

Wampfler, Philippe (2023): Schreiben mit KI-Tools, online unter: <https://tinyurl.com/zm4afkdd> (letzter Zugriff: 20.02.2025).